



---

Year: 2011

---

## **An efficient coding hypothesis links sparsity and selectivity of neural responses**

Blättler, F ; Hahnloser, Richard H R

**Abstract:** To what extent are sensory responses in the brain compatible with first-order principles? The efficient coding hypothesis projects that neurons use as few spikes as possible to faithfully represent natural stimuli. However, many sparsely firing neurons in higher brain areas seem to violate this hypothesis in that they respond more to familiar stimuli than to unfamiliar stimuli. We reconcile this discrepancy by showing that efficient sensory responses give rise to stimulus selectivity that depends on the stimulus-independent firing threshold and the balance between excitatory and inhibitory inputs. We construct a cost function that enforces minimal firing rates in model neurons by linearly punishing suprathreshold synaptic currents. By contrast, subthreshold currents are punished quadratically, which allows us to optimally reconstruct sensory inputs from elicited responses. We train synaptic currents on many renditions of a particular bird's own song (BOS) and few renditions of conspecific birds' songs (CONs). During training, model neurons develop a response selectivity with complex dependence on the firing threshold. At low thresholds, they fire densely and prefer CON and the reverse BOS (REV) over BOS. However, at high thresholds or when hyperpolarized, they fire sparsely and prefer BOS over REV and over CON. Based on this selectivity reversal, our model suggests that preference for a highly familiar stimulus corresponds to a high-threshold or strong-inhibition regime of an efficient coding strategy. Our findings apply to songbird mirror neurons, and in general, they suggest that the brain may be endowed with simple mechanisms to rapidly change selectivity of neural responses to focus sensory processing on either familiar or unfamiliar stimuli. In summary, we find support for the efficient coding hypothesis and provide new insights into the interplay between the sparsity and selectivity of neural responses.

DOI: <https://doi.org/10.1371/journal.pone.0025506>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-60598>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Blättler, F; Hahnloser, Richard H R (2011). An efficient coding hypothesis links sparsity and selectivity of neural responses. PLoS ONE, 6(10):e25506.

DOI: <https://doi.org/10.1371/journal.pone.0025506>

# An Efficient Coding Hypothesis Links Sparsity and Selectivity of Neural Responses

Florian Blättler, Richard H. R. Hahnloser\*

Institute of Neuroinformatics, University of Zurich/ETH Zurich, Zurich, Switzerland

## Abstract

To what extent are sensory responses in the brain compatible with first-order principles? The efficient coding hypothesis projects that neurons use as few spikes as possible to faithfully represent natural stimuli. However, many sparsely firing neurons in higher brain areas seem to violate this hypothesis in that they respond more to familiar stimuli than to nonfamiliar stimuli. We reconcile this discrepancy by showing that efficient sensory responses give rise to stimulus selectivity that depends on the stimulus-independent firing threshold and the balance between excitatory and inhibitory inputs. We construct a cost function that enforces minimal firing rates in model neurons by linearly punishing suprathreshold synaptic currents. By contrast, subthreshold currents are punished quadratically, which allows us to optimally reconstruct sensory inputs from elicited responses. We train synaptic currents on many renditions of a particular bird's own song (BOS) and few renditions of conspecific birds' songs (CONs). During training, model neurons develop a response selectivity with complex dependence on the firing threshold. At low thresholds, they fire densely and prefer CON and the reverse BOS (REV) over BOS. However, at high thresholds or when hyperpolarized, they fire sparsely and prefer BOS over REV and over CON. Based on this selectivity reversal, our model suggests that preference for a highly familiar stimulus corresponds to a high-threshold or strong-inhibition regime of an efficient coding strategy. Our findings apply to songbird mirror neurons, and in general, they suggest that the brain may be endowed with simple mechanisms to rapidly change selectivity of neural responses to focus sensory processing on either familiar or nonfamiliar stimuli. In summary, we find support for the efficient coding hypothesis and provide new insights into the interplay between the sparsity and selectivity of neural responses.

**Citation:** Blättler F, Hahnloser RHR (2011) An Efficient Coding Hypothesis Links Sparsity and Selectivity of Neural Responses. PLoS ONE 6(10): e25506. doi:10.1371/journal.pone.0025506

**Editor:** Stefan J. Kiebel, Max Planck Institute for Human Cognitive and Brain Sciences, Germany

**Received:** July 13, 2011; **Accepted:** September 5, 2011; **Published:** October 13, 2011

**Copyright:** © 2011 Blättler, Hahnloser. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** F.B. was supported by a TH grant of the ETH Zurich (TH-03 06-1), and R.H. was supported by the Swiss National Science Foundation (SNF). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: rich@ini.phys.ethz.ch

## Introduction

Brains of higher vertebrates analyze the sensory world in hierarchical networks. In lower sensory brain areas, neurons tend to respond to generic stimulus features, whereas in higher areas they typically respond to only very small subsets of natural stimuli [1,2]. Neural responses in lower areas are usually characterized by the stimulus property that correlates most strongly with spike responses, based on which neurons are termed feature detectors for that particular property. For neurons in higher brain areas, the concept of feature detector is often abandoned in favor of stimulus selectivity (assessed in terms of the stimulus that elicits the maximal response), e.g., neurons in higher auditory areas are selective for a particular birdsong [3], or, in visual areas, they are selective to the face of a particular person [4].

Feature tuning in sensory neurons can be explained by a neuronal strategy to efficiently encode natural stimulus ensembles, such as e.g. in simple cells of primary visual cortex or in auditory cells of the cochlear ganglion [5,6]. However, it is unclear whether (unsupervised) efficient coding principles can also explain complex response selectivity, as proposed in [7], in particular when selectivity applies to a behaviorally relevant stimulus such as the tutor song for a juvenile songbird. Naively, the selectivity for tutor song (TUT) or the bird's own song (BOS) that is commonly

observed in song-control brain areas [8,9] seems to violate the efficient coding hypothesis, according to which responses to frequent stimuli should be minimized [7,10], not maximized. Hence, it is an open question how complex selectivity (maximal response to BOS) can be reconciled with the sparse firing that often accompanies such selectivity.

The rationale of our work is that the firing sparsity of cells is governed by slow developmental mechanisms that construct efficient representations of natural stimulus statistics and by faster mechanisms influenced by the recent stimulus history and the state of the animal and its environment. In many cells, excitatory and inhibitory currents are balanced in proximity of the firing threshold, though currents can be dominated by inhibition, as for example in some sparsely firing neurons [11]. Also, shifts in the excitatory/inhibitory balance are commonly observed, for example as a function of stimulus intensity [12] or during sensory adaptation [13]. Mechanistically, the excitation-inhibition balance can be controlled by neuromodulatory mechanisms such as the serotonergic system [14], and, a link has been suggested to exist between the excitation-inhibition balance and attention [15]. We study the dependence of response selectivity in model neurons on fast shifts in the balance between excitatory and inhibitory inputs.

In our model, we reflect the asymmetry imposed by the firing threshold (supra- versus sub-threshold responses) by an asymmetric

cost on sub- and suprathreshold synaptic currents. We train neurons using a particular firing threshold and thereafter we explore the consequences of shifts in the excitatory/inhibitory balance by evaluating neural responses for an entire range of firing thresholds.

Our model is applicable to sensory systems in general but presented as a model of the auditory forebrain pathway of songbirds. This pathway extends over the nucleus ovoidalis to field L, and from there up to HVC [9,16,17], Figure 1A. A crucial function of this pathway is to subserve song learning, a process in which birds first memorize a template of a tutor's song (TUT) and then refine their vocalizations to gradually approximate the template [18]. During these two learning phases, neurons in the lower auditory brain area field L develop a broad response selectivity for natural sounds, in particular the songs of conspecific birds (CON) [19]. By contrast, neurons in higher areas develop selectivity for the TUT and the BOS in particular [20,21]. Preference for BOS over CON and REV is first observed in the caudal mesopallium (CM). BOS preference is even stronger in the nucleus interface of the nidopallium (Nif) and highest in the premotor area HVC [8,9]. Notable about this gradual increase in BOS selectivity along the main auditory pathway is a conjunctive increase in firing sparsity, illustrated by Nif neurons firing rather densely and HVC projection neurons bursting only once during a song or song motif [8,22].

## Results

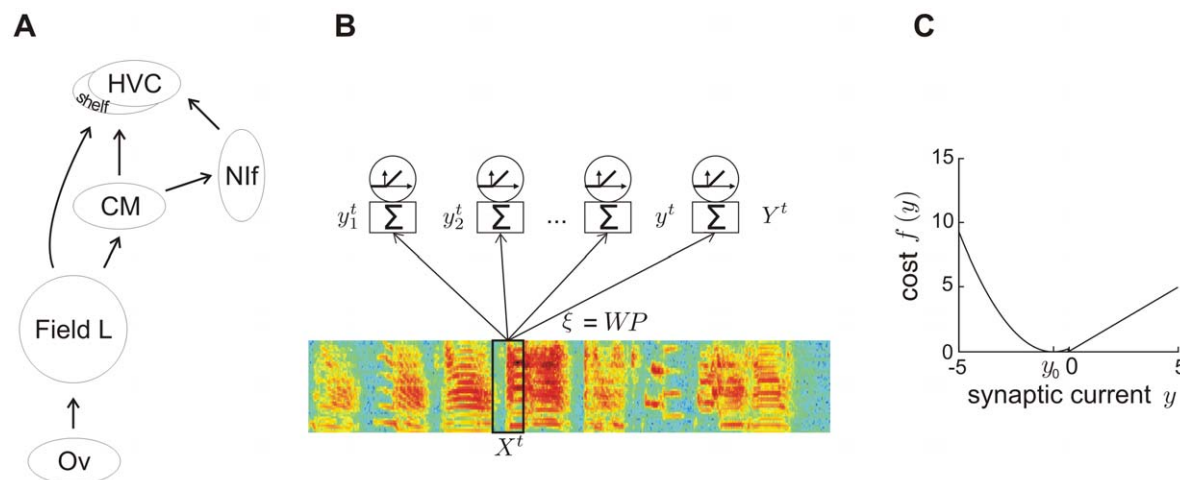
We model the auditory pathway as a feedforward network that receives (indirect) auditory input from the cochlea in the form of spectro-temporal sound patterns. These patterns are mean-subtracted, multiplied by synaptic weights, and summed, to result in the total synaptic current impinging onto neurons. Given this linear model we interpret the set of synaptic weights onto a neuron as its spectro-temporal receptive field (STRF). We devised an algorithm that optimizes synaptic weights for their propensity to decorrelate (whiten) and sparsify synaptic inputs: First, we

whitened cochlear inputs using a projection matrix  $P$  (principal component analysis, PCA), and then we applied a sparseness transformation using a matrix  $W$  that minimizes an asymmetric cost imposed on total synaptic currents, Figure 1B. This cost depends on the firing threshold  $\theta$ . Subthreshold synaptic currents are punished quadratically, whereas suprathreshold currents are punished linearly. The location of minimal quadratic cost is freely selectable using the parameter  $y_0$ , Figure 1C. For simplicity, in most simulations we neglected this parameter by setting  $y_0 = 0$  (we set  $y_0 \neq 0$  when our goal was to optimally reconstruct cochlear inputs from suprathreshold currents, in which case there is a better choice for  $y_0$  than zero, see Methods).

This model guarantees zero mean synaptic current and minimal variance (see Equation 8). Intuitively, to minimize the cost, frequent cochlear input patterns must elicit weak subthreshold currents (quadratic cost is smaller than linear cost). By contrast, rare cochlear input patterns must elicit strong suprathreshold currents (linear cost smaller than quadratic). We defined firing rates to be equal to suprathreshold synaptic currents, meaning that the linear cost of suprathreshold currents is in effect a cost on the average population firing rate. In the Methods we show that the quadratic cost of subthreshold currents is a cost on a reconstruction error associated with simple decoding of cochlear inputs from firing rates. Thus, the algorithm tries to maximally sparsify population responses without discarding any relevant sensory information. We minimized the cost function over training data consisting mostly of renditions of a particular zebra finch song (BOS) and a few CONs. After training, we evaluated neural responses for a wide range of firing thresholds (thresholds during training could deviate from thresholds during evaluation). Changes in firing thresholds modeled changes in the excitatory/inhibitory balance. Details are presented in the Methods.

## Spectro-temporal receptive fields (STRFs)

After training, neurons displayed a large diversity of STRFs. Typically, STRFs were patchy and had multiple adjacent inhibitory



**Figure 1. Efficient coding of cochlear spectrograms.** (A) Schematic of the auditory pathway in the songbird forebrain. Auditory input to the pallial field L is provided by the thalamic nucleus ovoidalis (Ov). From field L, auditory information is relayed to the caudal mesopallium (CM), and from there to the nucleus interface of the nidopallium (Nif) and to HVC. (B) Network model. At time  $t$ , the auditory input to the network is a 50-ms window  $X^t$  of the sound spectrogram. This input is multiplied by synaptic weights  $\xi = WP$  to result in total synaptic currents  $Y^t = (y_1^t, \dots, y_N^t)$  onto  $N$  neurons.  $P$  stands for whitening and dimensionality reduction (principal component analysis), and  $W$  stands for a sparseness transformation. Neural firing rates are given by rectified synaptic currents. (C) Cost function for a threshold  $\theta = 0$ . Subthreshold synaptic currents  $y$  are punished quadratically and suprathreshold currents are punished linearly. The parameter  $y_0$  defining the subthreshold current eliciting minimal cost is set to the expected subthreshold current  $y_E$ . Synaptic currents are reported in units of mean-subtracted standard deviations (z-scores). A threshold  $\theta = 0$  implies that suprathreshold currents are depolarizing (positive), whereas subthreshold currents are hyperpolarizing (negative).  $c = 1$ . doi:10.1371/journal.pone.0025506.g001

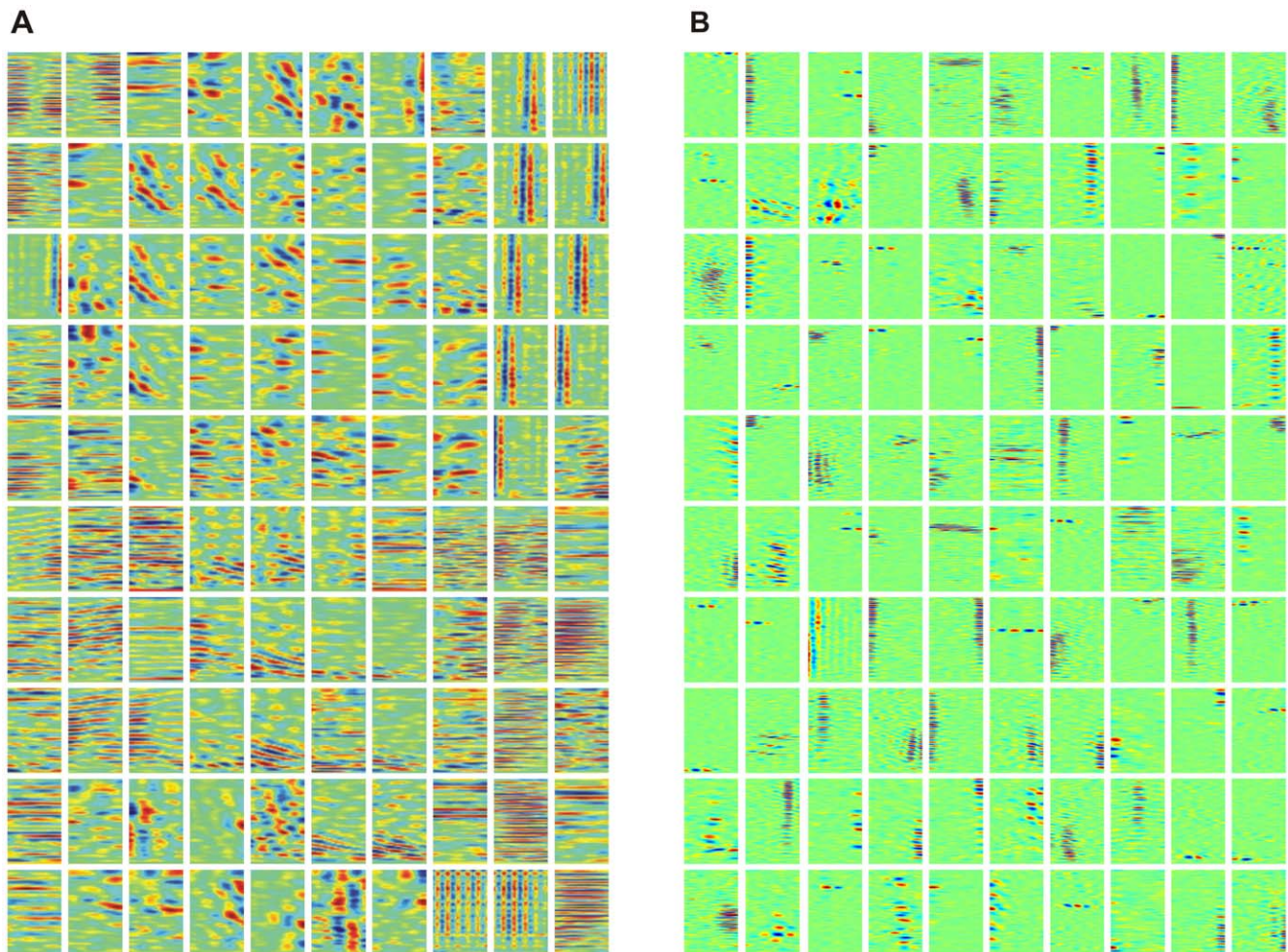


and excitatory spectral/temporal subfields. In many neurons, STRFs were regularly arranged into horizontal or vertical stripes (Figure 2A), similar to receptive fields in field-L neurons that encode elementary spectro-temporal sound features such as sound onsets or a particular sound pitch [23–26].

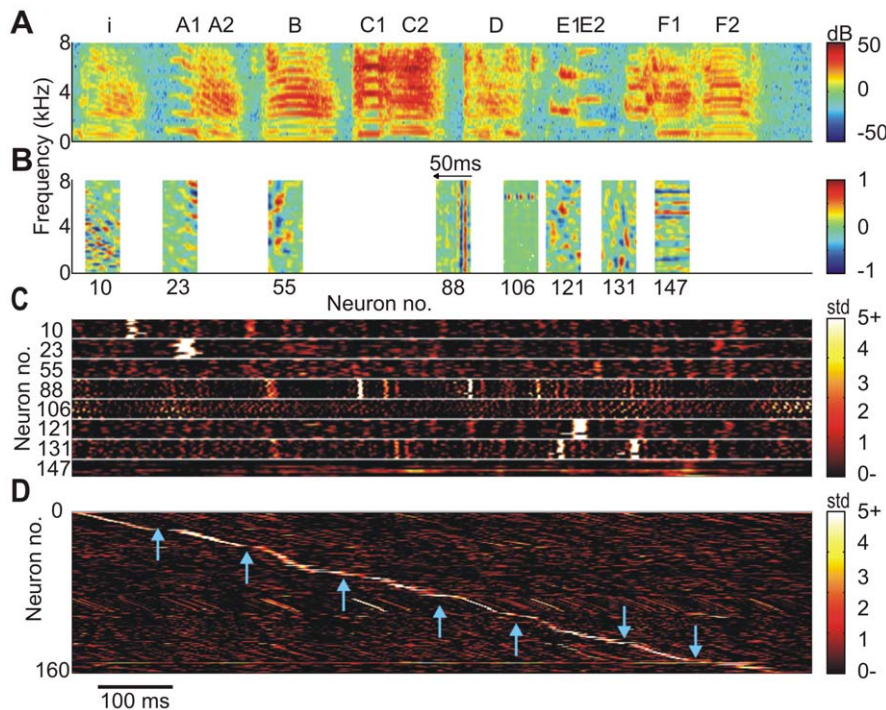
Model STRFs had excitatory and inhibitory subfields that together covered the entire spectro-temporal window of the STRF. Typically STRFs in field L are of considerably lower density in that they mostly possess only two or three subfields instead of more than six. We therefore explored whether STRFs in our model would be of lower density when we added a third term to the cost function, a term corresponding to a linear cost on absolute synaptic weights. We found indeed low-density STRFs if the parameter  $c_s$  weighing this third term exceeded roughly 0.1, Figure 2B. Density of STRFs could be controlled independently of the sparseness of model responses, explored next.

We explored the correspondence between STRFs and the stimulus features to which neurons responded most. In most cells, presentation of different BOS versions elicited reliable responses to specific song notes, Figure 3A–C. For example, the total synaptic

current in neuron 10 with a checkerboard-like STRF reliably peaked after the down sweep of the introductory note and to a lesser extent it also peaked at the offsets of some other syllables. Neuron 23 with a narrow and slanted STRF responded most strongly to the down-sweep of the harmonic stack in Syllable A1. The STRF of Neuron 88 had sharp vertical subfields, the synaptic current to this neuron peaked during rapid increases of sound intensity such as during the onsets of Syllables C and D. Another neuron with a vertically dominated STRF (Neuron 131) responded a few milliseconds after the onsets of Syllables E and F. This neuron was able to respond to different syllable onsets than Neuron 88 by virtue of its sensitivity to a low-frequency tone immediately followed by a high-frequency tone, which is a common characteristic of both Syllables E and F. Very particular was Neuron 106. Its receptive field and that of several other neurons were centered on a single frequency band close to 7 kHz. It turned out that this cell responded to electrical noise by which our recordings were affected; during BOS presentation, the total synaptic current to this cell was small and increased mainly during syllable gaps where no signal except the noise was present. The



**Figure 2. Spectral temporal receptive fields.** (A) Spectral temporal receptive fields (STRFs) of  $N = 100$  neurons, arranged by nearest-neighbor similarity (circular boundary conditions). Neurons tend to be either temporally tuned (vertical stripes, top right), spectrally tuned (horizontal stripes, middle rows), or display more complex spectro-temporal patterns. Spectral resolution is 172 Hz, offset between subsequent cochlear inputs is 1.5 ms. (B) STRFs obtained with a linear cost on synaptic weight magnitudes. The linear cost forces many synaptic weights to be close to zero (green), leading to low-density STRFs most of which contain a smaller number of excitatory and inhibitory subfields than in A. Interestingly, excitatory and inhibitory subfields tend to be close to each other and aligned horizontally or vertically, similar to observations in field L neurons. The 100 presented STRFs were randomly chosen out of the total 800.  $c_s = 0.2$ . doi:10.1371/journal.pone.0025506.g002



**Figure 3. Receptive fields and neurogram.** (A) Power spectrogram of a bird's own song (BOS). (B) STRFs  $\xi_i$  of eight representative neurons ( $i=10, 23, 55, 88, 106, 121, 131, 147$ ). The horizontal alignment of STRFs with the spectrogram in A is such that the trailing edges of the STRFs correspond to the respective peak times of synaptic currents. The temporal axis of the STRFs is inverted for better comparison with the BOS spectrogram. (C) Stack plot of synaptic currents of representative neurons in B in response to ten different versions of BOS, vertically aligned to A. (D) Neurogram of synaptic currents in response to the BOS in A. The  $N=160$  neurons are sorted according to the peak times of their synaptic currents. Fewer neurons display synaptic current peaks during syllable gaps (blue arrows) than during syllables.  
doi:10.1371/journal.pone.0025506.g003

structure of the sparsely checked STRF of Neuron 121 was particularly well adapted to Syllable E, the neuron responded almost exclusively during the transition between sub-Syllables E1 and E2. Finally, Neurons 55 and 147 did not show either robust or strong responses to BOS; a more thorough analysis revealed that they responded strongly to a CON in the training set. When we ordered all neurons by the time at which their synaptic currents peaked in response to a particular version of BOS, we found that the resulting stack plot exhibited a staircase-like shape: peak currents were widely distributed across cells with many more peaks seen during syllables than during syllable gaps, Figure 3D.

Some neurons were not just tuned to a particular syllable within the motif, but had even more specific tuning to syllable subsets of that syllable. Finches often produce harmonic stack syllables and can subtly vary the pitch of these syllables in a well-controlled and goal-directed manner [27,28]. When we trained a network of 196 neurons on the songs of a bird produced during an entire day, we found that synaptic currents of two neurons peaked during a harmonic-stack syllable produced by that bird (neurons with such harmonic-stack receptive fields have been illustrated in [26]). Interestingly, for any given stack syllable, the synaptic current in only one of the neurons peaked, but not in both, Figure 4. The two neurons divided the representation of that syllable between each other, one represented the high-pitch version of the stack, the other the low-pitch versions, Figure 4E. Hence, our algorithm is able to 'allocate' more than a single neuron to a song feature, depending on the extent of its variability.

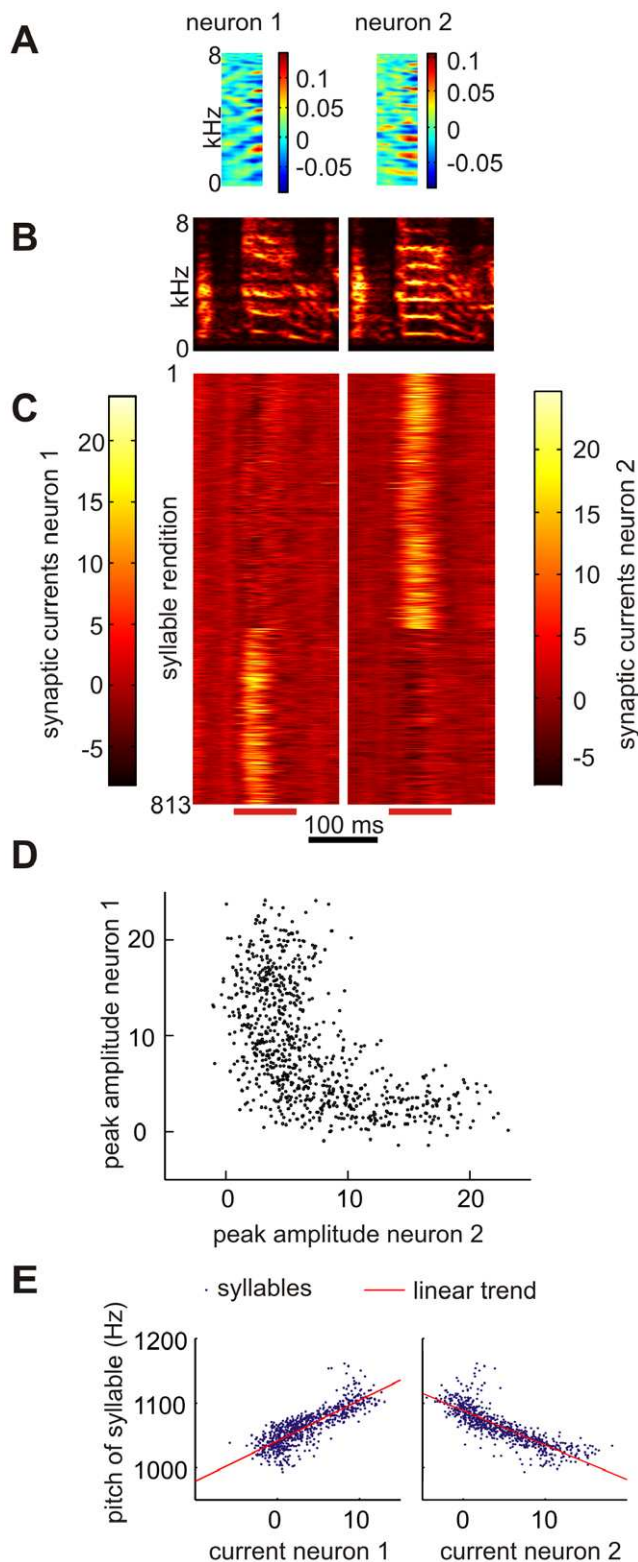
STRFs are optimal models of the linear part of neural responses. Can we recover the STRFs found by our algorithm in simulated neural responses which contain a threshold

nonlinearity? To this end, we estimated STRFs from nonlinear responses to birdsong stimuli for a range of firing thresholds  $\theta$ . We estimated STRFs using reverse correlation. Mostly, we found strong resemblance between estimated and actual STRFs, Figure 5. Strong resemblance was seen in all cases in which the correlation coefficient  $cc$  between estimated and actual responses was above 0.1, i.e. in cases in which the linear model was reasonably good. Moreover, estimated STRFs did not change much with increasing firing threshold  $\theta$ , except that with increasing  $\theta$  the STRFs had a small tendency to extend over larger time-frequency regions (Figure 5B) than the original low-density STRFs (Figure 5A). We found similarly satisfying results when estimating high-density STRFs (not shown). Hence, estimated STRFs were quite insensitive to the nonlinearity and to changes in firing threshold.

### Distribution of synaptic currents

The distribution of total synaptic currents over all neurons and over all training stimuli was highly asymmetric and contained many positive but few negative outliers, Figure 6A. The distribution of BOS-evoked currents could be reasonably well approximated by a unit Gaussian on the negative side and a long-tail exponential on the positive side. This combined Gaussian-exponential behavior follows from the fact that minimization of the quadratic-linear cost function is equivalent to locally maximizing the log-likelihood of a Gaussian model density below the threshold and of an exponential model density above the threshold, under the global restriction of zero mean and fixed variance. Interestingly, large synaptic currents were mostly elicited by the BOS rather than by other stimuli, illustrating that neurons





**Figure 4. Neurons encode behavioral variability, for example song pitch.** (A) Two receptive fields formed by training a network on all songs produced by a bird on a single day. (B) Spectrograms of a song syllable containing a harmonic stack. The left version has median pitch 1024 Hz, the right version 1138 Hz. (C) Stack plot of synaptic currents in the two neurons elicited by 813 syllable renditions. The stack plots have been sorted identically to reveal that for a given syllable rendition either the left or right neuron exhibits a peak in synaptic current, but not both.

Peaks in synaptic currents are computed in intervals indicated by red bars on the bottom. (D) Scatter plot of peak synaptic currents in the two neurons. The distribution is sparse ('L-shaped'). (E) Median synaptic current in same intervals versus median song pitch of the harmonic stack. The two neurons are detectors of low and high pitch versions of the stack, respectively. Red and blue lines are linear regressions (Neuron 1:  $R^2=0.68$ ,  $p<10^{-170}$ , Neuron 2:  $R^2=0.69$ ,  $p<10^{-185}$ ,  $N=196$ ,  $y_0=0$ ).

doi:10.1371/journal.pone.0025506.g004

were best tuned to the features of the most prominent stimulus in the training set, which was the BOS. The same finding was true for low-density STRFs (when  $c_s > 0.1$  instead of zero): BOS-elicited currents exhibited a heavier positive tail than currents elicited by CON and REV (Figure 6B). Hence, model responses were robustly tuned for the BOS.

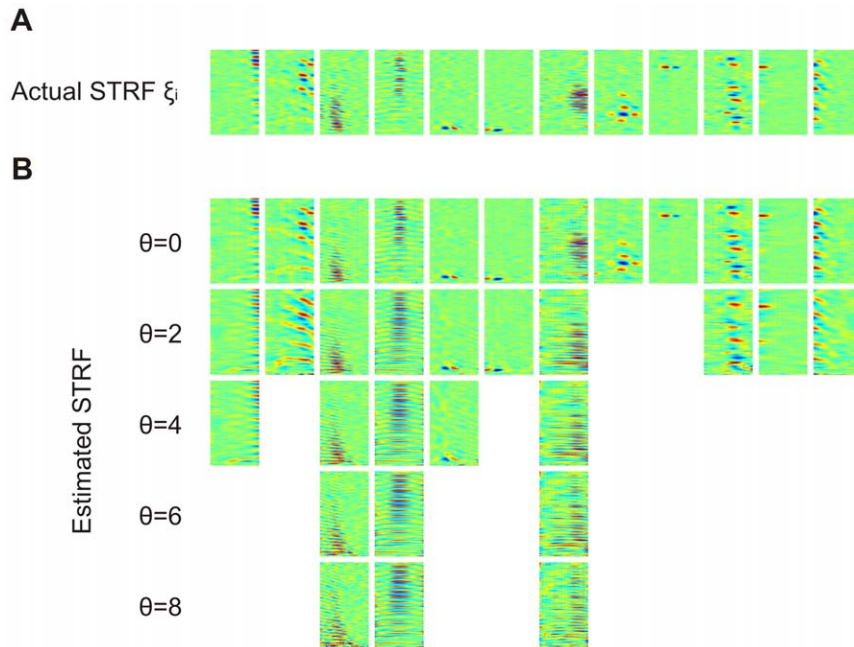
We also explored the influence of other model parameters such as  $y_0$ , which sets the location of minimal quadratic cost. When we changed  $y_0$  to nonzero values different from the firing threshold  $\theta$  during training, we found that BOS tuning of synaptic currents was qualitatively unchanged. The only effect of changing  $y_0$  was to slightly increase the distribution of synaptic currents around  $y_0$  (where cost is minimal) and to slightly decrease it around  $\theta$  (not shown).

### Distribution of firing rates

We computed firing rates in model neurons by thresholding total synaptic currents. A recent analysis of sparsely firing cells in primary auditory cortex of unanesthetized rats has revealed that both background and stimulus-evoked firing rates are well fit by log-normal distributions [29]. We speculated that log-normal firing may be a corollary of efficient coding that could be reproduced in simulations. We inspected the distributions of firing rates across all neurons for all BOS and CON stimuli. Indeed, we found that the density of firing rates was best fit by a log-normal distribution, which was especially true for low firing thresholds, Figure 7. Note that recently published firing rate distributions of field L neurons in zebra finch were fit by Gamma distributions [30]. However, the published data suggests that a fit with a log-normal distribution should be equally good.

### Independence of neural responses

Training the network increased the independence of neural responses. If responses were perfectly independent among neurons, the size distribution of coactive neurons would be binomial (the size distribution is the probability that a given number of neurons fire synchronously). The sole parameter of this binomial model is the single-neuron firing density that we estimated in terms of the fraction of suprathreshold events observed in the entire neuron population and for all training stimuli. In comparison to this binomial model, the observed size distribution elicited by whitened cochlear inputs was substantially wider, illustrating strong firing dependencies. The sparseness transformation significantly narrowed the observed size distribution towards the binomial case, Figure 8A. This increase of independence (decrease in Kullback-Leibler divergence to the binomial model) was true for both high and low firing densities, and true for nearly all firing thresholds tested, Figure 8B, revealing that the sparseness transformation is a robust mechanism to increase independence of neural responses. Qualitatively, this behavior of the network to render responses more independent applied to all firing thresholds used during training (we tested thresholds up to  $\theta=5$ ).



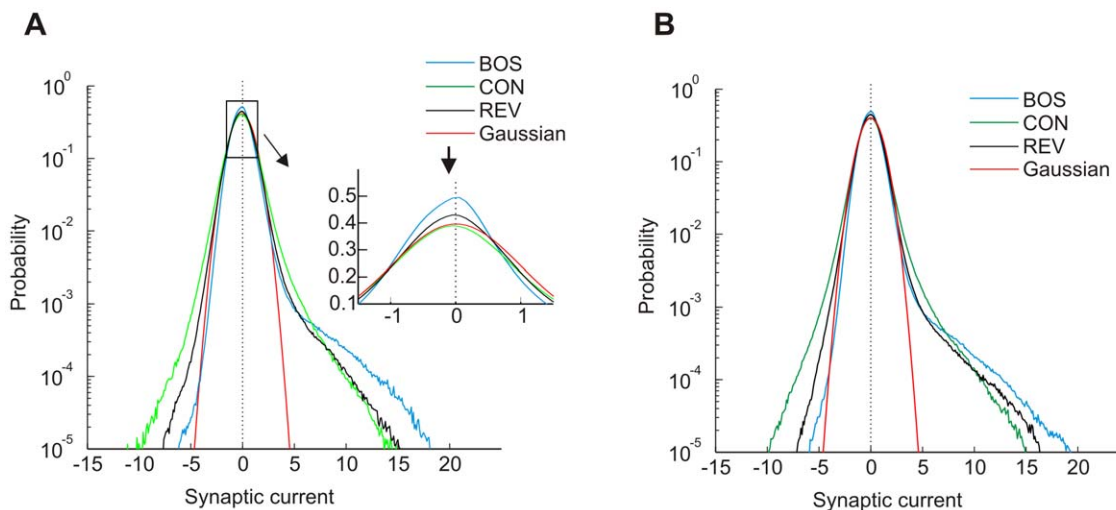
**Figure 5. Actual STRFs and STRFs estimated using reverse correlation.** (A) A selection of twelve STRFs  $\xi_i$  obtained after convergence of the algorithm ( $N = 800$ ). (B) Estimated STRFs (reverse correlation) based on the predicted firing rates  $r_i^r$ . Shown are only estimated STRFs for neurons associated with a correlation coefficient  $cc$  between predicted and actual firing rates of  $cc > 0.1$ .  $\mu = 1$ .  $c_s = 0.2$ . doi:10.1371/journal.pone.0025506.g005

### Decoded firing and reconstructed spectrograms

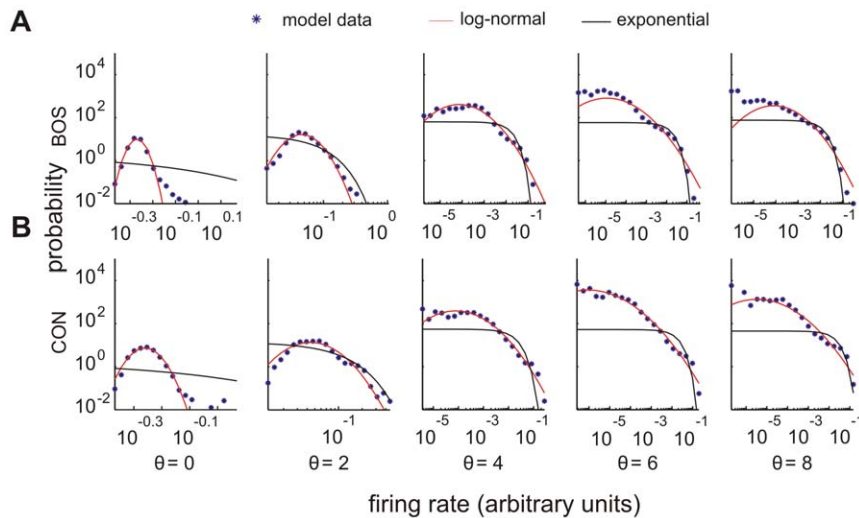
Our network allowed us to decode firing rates and reconstruct the spectro-temporal sound patterns that elicited them, using the pseudo-inverse of the sparseness transformation  $W$  (see Methods). We evaluated the reconstructions for various firing thresholds (after training at a fixed threshold of zero), Figure 9A-C. For a threshold of zero, neurons produced dense firing patterns in response to BOS, with roughly 50 percent of neurons active at any

time, Figure 9D. The percent active neurons decreased from 20% for  $\theta = 1$ , to 1–2% for  $\theta = 3$ , and down to 0.4% for  $\theta = 5$ . At thresholds higher than roughly three, reconstruction errors associated with non-BOS stimuli were often due to missed syllables because none of the neurons fired in response to these syllables.

In all cases, reconstruction errors increased with increasing firing threshold in a monotonic manner, Figure 9E. For a given



**Figure 6. Probability density of total synaptic currents.** (A) The probability density of total synaptic currents  $y$  averaged over all neurons has a heavy tail on the positive side. Shown are the densities for BOS (blue), CON (green), and REV (black). Near zero synaptic currents, the curves are approximately unit Gaussian (red), though their excessive peaks are slightly shifted to the negative side (inset, arrow). The curves cross each other such that large positive synaptic currents are preferentially elicited by the BOS and small positive currents by REV and CON.  $N = 400$ ,  $y_0 = y_{E-}$ . (B) The distributions of synaptic currents for sparse STRFs (Figure 2B) are qualitatively similar to (A). The only noticeable difference is that the distribution for REV is closer to BOS, reflecting a lower selectivity for temporal order.  $c_s = 0.2$ . doi:10.1371/journal.pone.0025506.g006



**Figure 7. Probability densities of mean firing rates.** Mean firing rates in response to (A) a BOS stimulus and (B) a CON stimulus. For each cell we computed the mean firing rate to one stimulus trial. Our simulation data (blue asterisks) are better fit by log-normal densities (red) than by exponential densities (black). Firing rates are plotted in arbitrary units. Fit parameters for log-normal densities were determined by the mean and variance of logarithmic firing rates, and for exponential densities they were determined by the mean firing rates. Thresholds varied from  $\theta=0$  to  $\theta=8$ . Noise amplitude  $k=1$ .

doi:10.1371/journal.pone.0025506.g007

threshold, reconstructions from sparseness-transformed cochlear inputs were much better than reconstructions from merely whitened inputs. This superiority was true even though for thresholds up to approximately 1.3, mean firing rates were lower for sparseness-transformed inputs than for merely whitened inputs.

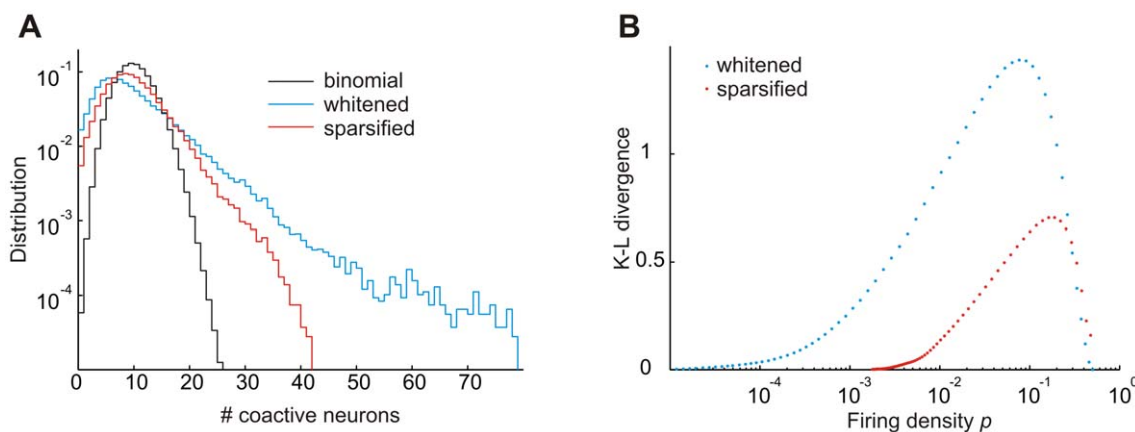
Moreover, for a given threshold, reconstruction errors tended to be larger for the BOS played back in reverse (REV) than for BOS or CON, illustrating that reconstructions were optimized for stimulus ensembles experienced during training but not for novel ensembles. In the Methods we show that the reconstruction error is approximately equal to a term that grows not only with the number of subthreshold events, but also with their variance; hence, stimuli that induce narrow subthreshold distributions (such as the BOS, Figure 6) lead to smaller reconstruction errors.

### Smart noise suppression by selective neuron exclusion

As illustrated in Figure 10, five STRFs were encoding electrical noise elicited by the computer monitor. When we omitted these five neurons for BOS reconstructions, we were able to effectively suppress the monitor noise. To demonstrate the effectiveness of this smart noise suppression, we iteratively estimated the sound waveform associated with the reconstructed BOS [31]. The original and the noise-suppressed waveforms are available as supporting information files Audio S1 and S2 in wav format.

### Song selectivity is linked to firing sparseness

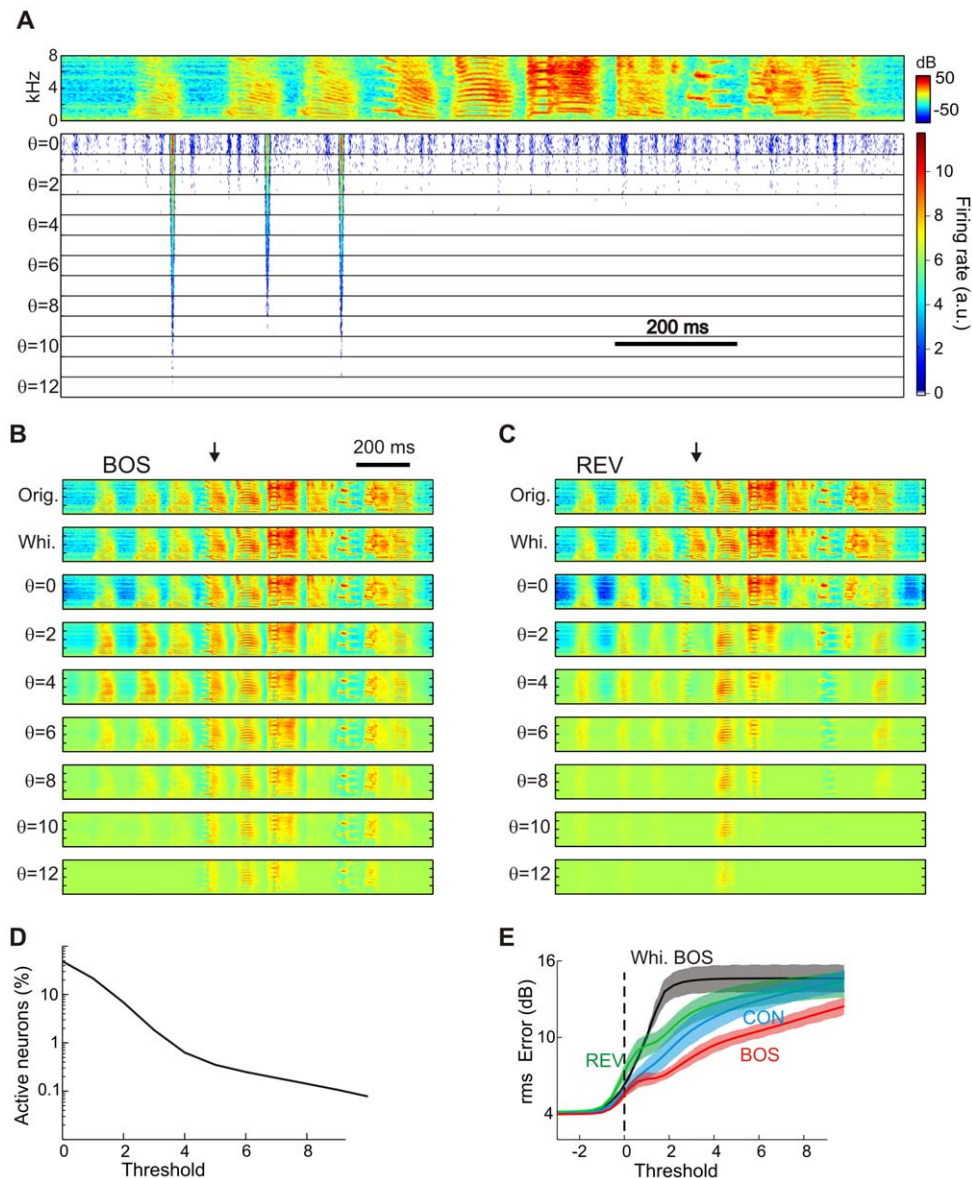
We explored the response selectivity of model neurons using the psychophysical  $d'$  measure [32] that is routinely applied in birdsong studies. According to this measure, the selectivity for a



**Figure 8. Sparsification reduces firing dependences.** (A) The sparseness transformation renders the size distribution of coactive neuron groups (whitening+sparseness) closer to binomial. The probability  $p$  of the binomial distribution (that a neuron is active per unit time) was estimated in terms of the firing density (the fraction of suprathreshold events over all neurons and training stimuli). Probabilities  $p$  were nearly identical for whitening and whitening+sparseness when  $\theta=2$  ( $\theta=0$  and  $y_0=y_{E-}$  during learning). (B) The Kullback–Leibler divergence between size distributions is smaller when comparing the whitening+sparseness model to the binomial model than when comparing the whitening model to the binomial model, for nearly all firing densities tested.

doi:10.1371/journal.pone.0025506.g008





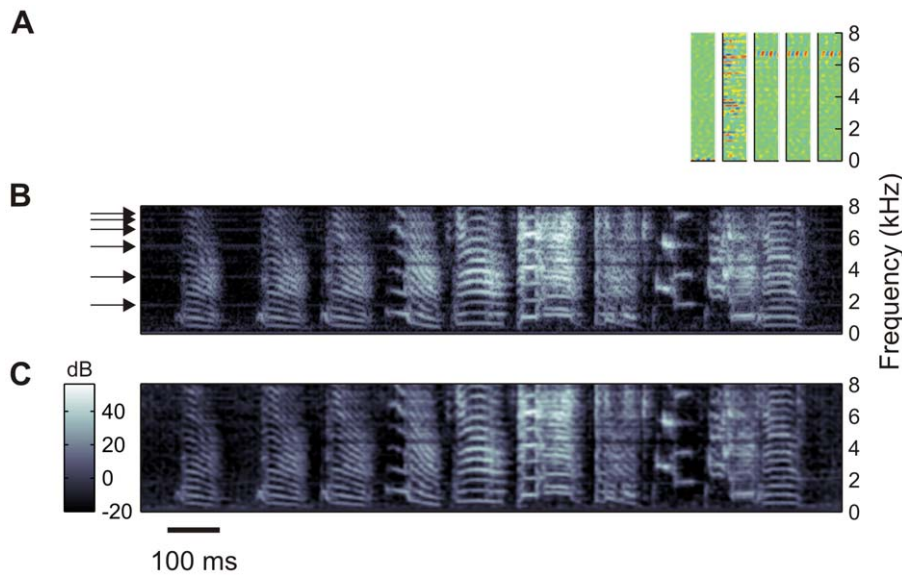
**Figure 9. Reconstructing the cochlear spectrograms from firing rates.** (A) Firing-rate of one example neuron in response to BOS for increasing firing thresholds ( $\theta = 0$  to 12). The BOS spectrogram is shown on top. This neuron is tuned to a feature present in introductory notes and responds to it up to thresholds higher than seven. For each threshold, ten different responses are plotted, corresponding to ten different instantiations of synaptic noise.  $k = 1$ . (B) The reconstruction of a BOS spectrogram (orig., top) using all neurons, based on a firing threshold of minus infinity (whi., 2nd from top) is fairly complete with little information loss (arising from dimensionality reduction). With increasing thresholds (below), more and more syllables are lost in the reconstruction, but the reconstructed spectro-temporal patterns remain clearly recognizable. The arrow points to a down-sweep syllable. (C) Reconstructions of REV (flipped horizontally for comparison with B) are worse than reconstructions of BOS at the same threshold; for example the down-sweep syllable is not well reconstructed (arrow), presumably because zebra finches produce almost no up-sweeps. (D) The fraction of active neurons (averaged over all BOS stimuli) decreases with increasing threshold such that at  $\theta = 3$  about 1% of neurons are active on average. This fraction decreases to 0.1% at about  $\theta = 9$ . (E) The reconstruction errors averaged over different stimulus ensembles are monotonic functions of the firing threshold. For a given positive threshold, reconstruction errors increase from BOS to CON to REV.  $N = 400$  threshold-linear neurons.

doi:10.1371/journal.pone.0025506.g009

stimulus over another is given by the difference in mean firing rates elicited by these stimuli, normalized by their standard deviations (see Methods). We assessed the selectivity of neurons to BOS versus matched spectro-temporal stimuli such as CON and REV. Variability of responses to a fixed stimulus was generated by a white-noise current source (see Methods).

We found a wide range of selectivity behaviors. Many neurons responded more strongly to CON than to BOS, but this CON

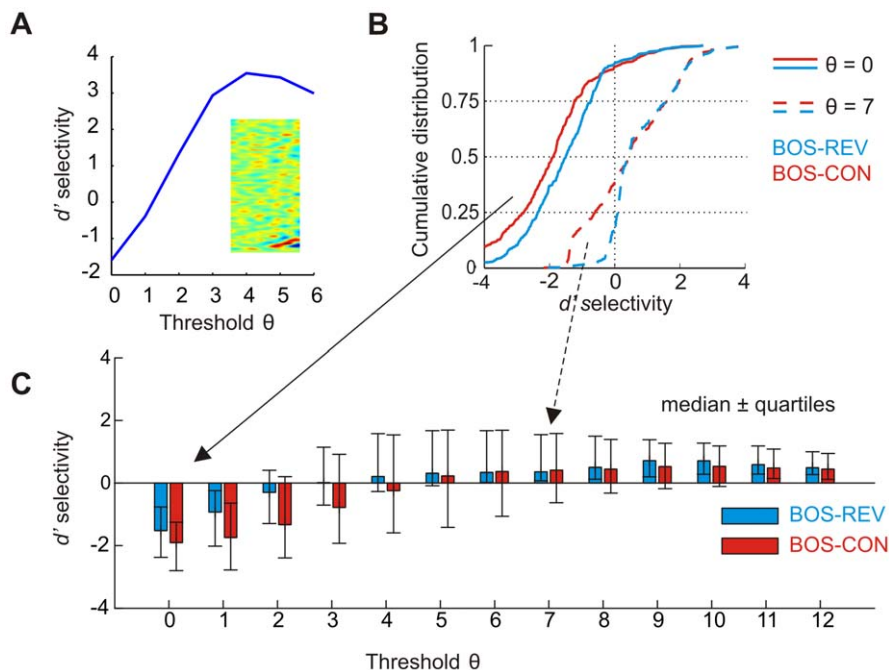
preference often reversed to BOS preference at high firing thresholds, Figure 11B. For a firing threshold of zero, the median  $d'$  selectivity for the BOS was negative across the population, both with respect to REV and to CON, Figure 11B. Hence, at this low threshold, the majority of neurons preferred REV and CON over BOS. BOS anti-preference remained true for a range of firing thresholds  $\theta$  up to three. However, all of the median and mean BOS-REV and BOS-CON selectivities became positive at



**Figure 10. Smart suppression of electrical noise affecting the recordings.** (A) The STRFs of five neurons that encoded monitor noise. (B) Original BOS spectrogram. The noise is manifest as gray horizontal bands (black arrows) during syllable gaps. (C) Reconstruction of the BOS (Equation 14) from elicited responses in the network. The thresholds of all neurons were set to  $\theta = -\infty$ , with exception of the five neurons in A in which the thresholds were set to  $\theta = +\infty$ . The monitor noise has vanished in the reconstructions, without affecting the birdsong signal.  
doi:10.1371/journal.pone.0025506.g010

thresholds  $\theta \geq 5$ , Figure 11C. Thus, the response selectivity of the network was non trivial in that it reversed at higher thresholds. Overall, BOS preference at high firing thresholds was seen both

for analog models (firing rates are defined as thresholded synaptic currents) and for binary models (firing rates are defined as binarized synaptic currents). From the point of view of stimulus



**Figure 11.  $d'$  selectivity for BOS reverses at high firing thresholds.** (A) Example model neuron with reversing BOS-CON selectivity. This neuron's STRF (inlay) codes for an up-sweep from 500 to 800 Hz over 20 ms. The resulting BOS-CON  $d'$  selectivity is negative for low thresholds  $\theta = 0, 1$  and turns positive for thresholds  $\theta \geq 2$ . (B) Example cumulative distributions of BOS-REV (blue) and BOS-CON (red)  $d'$  selectivities across  $N = 400$  neurons for  $\theta = 0$  (solid lines) and  $\theta = 7$  (dashed lines). For  $\theta = 0$  the selectivities are biased towards negative values, whereas for  $\theta = 7$  the distributions are biased towards positive values. (C) Bar plot summarizing BOS-REV (blue) and BOS-CON (red) selectivities for a wide range of firing thresholds. The colored bars indicate the median  $d'$  selectivity and the error bars delimit the first and third quartiles. Selectivity reverses at around  $\theta = 3$  (REV) and  $\theta = 5$  (CON).  
doi:10.1371/journal.pone.0025506.g011

selectivity, the low-threshold regime of our network is a model of densely firing field-L neurons and the high-threshold regime is a model of sparsely firing HVC neurons.

Our training set contained many versions of two different CONs. For BOS-CON selectivity reversal it did not matter whether selectivity was tested on three novel CONs as in Figure 11, or on twelve novel CONs, or on the two trained CONs, because for all these cases the median and mean BOS selectivities reversed at around  $\theta=3-4$ . However, when CONs from many more birds ( $>20$ ) were in the training set, then the median BOS-CON selectivity became positive only at very high thresholds ( $\theta \geq 8$  for 22 CONs), whereas the mean selectivity became positive already at  $\theta \geq 3$ . Thus, when responses to many different songs are sparsified, then increasing numbers of neurons develop a feature preference that best matches a CON in the training set rather than the BOS; however, this match is not particularly good as illustrated by BOS that is preferred on average already at relatively low thresholds.

For any given threshold  $\theta$ , the median  $d'$  selectivity (be it positive or negative) depends on the noise level. When increasing the noise level, the median  $d'$  selectivity goes toward zero, and, when decreasing the noise level, the median  $d'$  selectivity diverges from zero.  $d'$  magnitudes are also influenced by the number of different song renditions used to probe selectivity. When selectivity is probed with a single BOS and a single CON file and noise is small,  $d'$  values can become arbitrarily large. Hence, our model allows for close to arbitrary scaling of  $d'$  values by manipulating the intrinsic noise.

We tested the dependence of BOS selectivity on the temporal summation window and found that our results did not depend critically on STRF width. Model STRFs were 50 ms wide. For 100-ms wide STRFs, BOS-CON and BOS-REV selectivities reversed at around  $\theta=2-3$ ; and, for 25-ms wide STRFs, selectivity reversal was seen at around  $\theta=4$ . Hence, with increasing temporal summation window, selectivity reversal was seen at lower thresholds. Note that 25-ms STRFs are much shorter than estimated integration times in BOS-selective neurons [23,33].

We also tested the effect of neuron number on  $d'$  selectivity. Doubling that number from  $N=400$  to  $N=800$ , or reducing it to  $N=200$  or  $N=100$  preserved selectivity reversal in the range  $\theta=2-5$ , for all CON ensembles tested and for both firing-rate models. Also, we found that the value of the threshold  $\theta$  during learning has little influence on selectivity reversal after learning. For  $\theta=0$ ,  $\theta=2$ , and  $\theta=5$  during learning, selectivity for BOS reversed to positive values at around  $\theta=3-4$  after learning. In

summary, selectivity reversal at high thresholds was very robust and did not depend on model details.

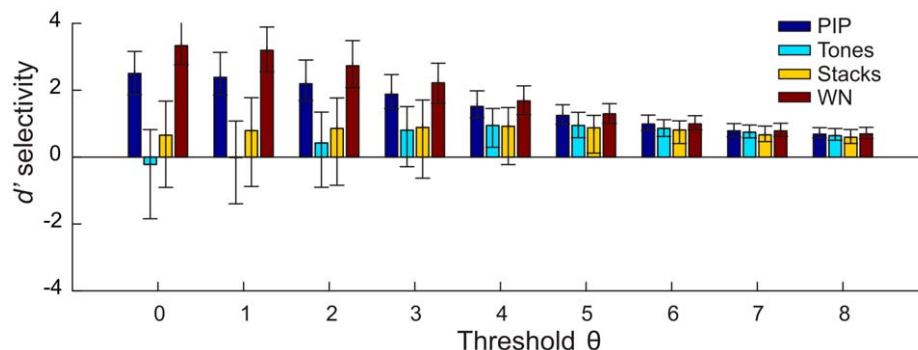
We assessed whether our model neurons preferred CON over artificial stimuli, as has been reported in field L [34,35]. We found high median selectivity for CON versus tone pips, tone stacks (ripples), and white noise (Figure 12). This CON preference was true for all thresholds  $\theta \geq 0$  examined. Solely tones (sparse colored noise) were preferred over CON for thresholds up to  $\theta=1$ . For higher thresholds  $\theta \geq 2$ , CON-tones selectivity reversed and CON was strongly preferred.

### Comparison with an ICA algorithm

The key element in our model seemed to be the positive tail of synaptic currents. Because this tail has a non-Gaussian shape, our model can be seen as part of a broader class of independent-component analysis (ICA) algorithms that extract maximally non-Gaussian components from data [36]. To test whether BOS preference at high thresholds arises also in other ICA algorithms, we trained an identical network using the classical ICA algorithm by Bell and Sejnowski (results not shown) [37]. In this algorithm, as in most similar algorithms, the final distribution of synaptic currents is symmetric, with heavy tails on both sides. For this reason we applied the firing threshold to absolute synaptic currents (see Methods). We computed BOS selectivities for different firing thresholds  $\theta$  and found that BOS-CON and BOS-REV selectivities reversed as in our model, but at higher thresholds: The median and mean BOS-REV selectivities reversed at  $\theta=4$ , whereas the mean and median BOS-CON selectivities reversed at around  $\theta=7$ . Thus, the emergence of BOS preference in the ultrasparse regime of simple networks did not depend on how efficient coding was enforced, but appears to represent a generic consequence of non-Gaussian statistics and the choice of the training set.

### Simulation of a two-layer network

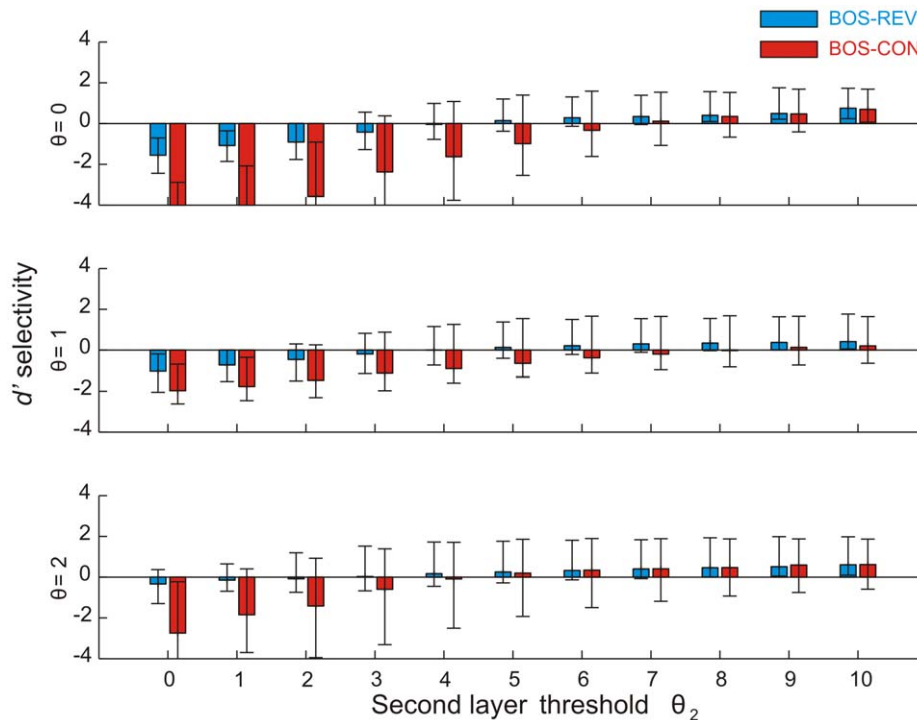
To account for the layered architecture of the auditory pathway, we also explored a two-layer network, in which the second layer was trained on thresholded first-layer outputs. In simulations, first-layer outputs were first summed over consecutive time bins (to extend receptive field widths in the second layer) and were then subjected to whitening and sparseness transformations (as we did for the first layer, see Figure 13). In simulated networks in which the first layer was a universal encoder (small  $\theta$ ), we found that response selectivity in the second layer reversed at high firing thresholds in preference of BOS. Hence, our high-threshold model



**Figure 12. Selectivity for CON vs. different artificial stimuli.** Depicted are median selectivities  $\pm$  quartiles. PIP = tone-pip stimuli, WN = white noise.

doi:10.1371/journal.pone.0025506.g012





**Figure 13. Selectivity in two layers.** Median and quartile selectivities in a network of two layers for various firing thresholds  $\theta$  in the first layer and  $\theta_2$  in the second layer. In each simulation, second-layer responses were evaluated using the first-layer threshold applied during training. As can be seen, BOS preference in the second layer is restricted to the high-sparseness regime there (right part of the three subplots). doi:10.1371/journal.pone.0025506.g013

of response selectivity in HVC can also be expressed in an architecture that is consistent with the feedforward organization of the auditory pathway.

## Discussion

Our model offers an understanding of auditory response selectivity based on an efficient coding hypothesis. We briefly outline the computational roots of the network we described, compare neural responses to those of songbird auditory neurons, and discuss possible consequences of efficient synaptic currents for sensory and motor processing.

### ICA algorithms

Our model falls into the broad category of ICA and sparse coding algorithms which try to linearly transform given (sensory) inputs into independent outputs (e.g., synaptic currents). In most ICA algorithms, independence of outputs is enforced by a symmetric cost such as kurtosis or entropy [37–39]. Because of this symmetry, most ICA algorithms fail to account for the asymmetry imposed by the spike threshold. By contrast, our algorithm explicitly includes a rectification nonlinearity which truncates as little information as possible because we minimize an approximate error of reconstructed spectrograms. Among the ICA algorithms that produce asymmetrically distributed outputs with a heavy tail, ours is most closely related to non-negative ICA [40], although in applications we found non-negative ICA to suffer from the problem that positive synaptic currents are completely unconstrained, which precludes interpretations of their magnitudes. A non-negative sparse-coding algorithm with a cost function similar to ours has been described [41], but imposes some tighter restrictions on the mixing matrix  $J$  (the inverse of  $W$ ) and has the

undesirable property that outputs  $y_i^j$  cannot be computed in a single forward pass but require an iterative optimization procedure. Last but not least, our algorithm is different from nonnegative matrix factorization (NMF) [42], because NMF does not allow for STRFs with inhibitory subfields.

### Song selective neurons in the auditory forebrain

In our model, the prominence of BOS in the training set and the shape of our cost function conjunctively forced neurons to display minimal suprathreshold synaptic currents to BOS (on average). These minimal responses explain why neurons preferred CON over BOS at moderately low firing thresholds, very much like field-L neurons do. At higher thresholds, neurons responded to specific BOS features more than to other features. In the high-threshold regime, model neurons were BOS selective, and during BOS presentation they fired sparsely and were hyperpolarized on average, all very much like HVC mirror neurons that project to Area X (HVC<sub>X</sub> neurons): HVC<sub>X</sub> neurons are hyperpolarized by playback of the BOS and produce a high frequency burst in response to a very specific song feature [11,22]. However, at high thresholds, our model network did no longer function as a universal encoder of the auditory environment. Although BOS reconstructions worsened very gracefully with increasing threshold, the neural representation of non-preferred stimuli degraded rapidly. This observation recommends the high-threshold regime only for specialized auditory areas such as HVC and the low-threshold regime for lower auditory areas such as field L that respond to a large variety of sounds.

Can selectivity for the TUT also be seen as a corollary of an efficient coding hypothesis? HVC neurons in juveniles tend to prefer TUT over most other stimuli including the BOS [21], whereas in adults this selectivity reverses such that HVC neurons

tend to prefer BOS over TUT [21,43]. Taken together, these data could be reproduced by our high-threshold model of HVC if TUT originally were the more prominent stimulus than BOS, but then BOS takes over as the most prominent stimulus.

Our model may explain response properties in many higher auditory brain areas of songbirds including CM: medial CM responses are shaped by auditory memories and cells typically respond more to familiar than to unfamiliar songs [44], in analogy to BOS preference seen in HVC. Interestingly, selective CM cells have lower spontaneous firing rates than non-selective cells, in agreement with the high-threshold regime of our model.

### Smart noise reduction

From an engineering perspective, one computational benefit of the sparseness transformation in our model is smart noise reduction. Using a high firing threshold during reconstruction, sounds to which cells have not been exposed during training can be effectively filtered out. More interestingly, by omitting certain neurons during reconstruction (e.g., by setting their firing thresholds to infinity), undesirable signals encountered in the training set can be conveniently suppressed. For example, by excluding the five neurons that encoded high-frequency noise (e.g. Neuron 106), BOS could be efficiently cleaned from that noise (see Audio S1 and S2). Of course the brain may make use of such smart noise reduction without ever explicitly having to reconstruct the original input; for example, feature-based attentional inputs may selectively suppress the firing in some neurons to constrain downstream processing to only relevant sensory features. Although such selective suppression has not been found yet in songbirds, birds may possess attentional selection mechanisms because they can detect subtle acoustic features and adapt their songs when negatively reinforced [27].

### Song feature analysis

Our algorithm offers a powerful method for bioacoustic signal analysis. The diversity of STRFs in our model is well matched with the behavioral richness of birdsong. Stereotyped syllables can be readily detected because they are represented essentially by a single STRF, whereas more variable syllables such as harmonic stacks may be associated with multiple STRFs as in Figure 4. Hence, the number of STRFs allocated to a particular syllable or sub-syllable may reflect its variability. The level of song analysis (detailed vs coarse) can be controlled by the number of neurons in the network. We can imagine uses of our algorithm for detecting particular song variants (such as high-pitched versions of harmonic stacks), or for identifying similar notes within different syllables, etc. Last but not least, because outputs of the network are more independent than its inputs, our algorithm may be well suited for applications of blind source separation.

### STRFs and their relation to spike responses

Our study is not the first to examine sparse coding of birdsong. Greene et al. applied a popular sparse coding algorithm to compute optimal linear kernels on large numbers of birdsong spectrograms [45]. They evaluated model output in terms of receptive field shapes and found that a stronger sparseness prior during training led to stronger resemblance of model STRFs with STRFs in field L. The STRFs in our similar model also qualitatively resembled STRFs in field L. Without density prior, model STRFs were denser in spectral and temporal modulations than field L STRFs. However, we showed there is no principled discrepancy because by using a suitable density prior, we were able to modulate the STRF density almost arbitrarily (Figure 2B), implying that our model is amenable to fitting a large variety of

experimental STRFs. Most importantly, our work suggests that neural firing may constitute a better model read out than receptive fields, because neural firing takes nonlinearities into account (such as the firing threshold), whereas receptive fields are linear and often poor descriptions of spike data. For example, we found that as a function of the firing threshold, our model, despite its fixed underlying STRFs, was able to reproduce qualitatively different responses as seen in field L and HVC. Hence, a simple STRF may be far from ideal as a characterization of neural firing, because it may be associated with a diverse range of response behaviors.

### Applicability to other sensory modalities

Our findings may have relevance for the encoding of sensory modalities other than audition, including olfaction. In mammals, strong odorant-selective responses arise immediately downstream of primary sensory inputs [46]. Though the neural mechanisms of this selectivity remain to be studied, in insects the mechanisms giving rise to sparse odor representations have been well characterized. The sparse odor representation in Kenyon cells arises from synchronized excitatory inputs mediated by densely firing projection neurons in the antennal lobe and by nonspecific inhibitory inputs from lateral horn interneurons that in essence set a high firing threshold to Kenyon cells [47,48]. The Kenyon cell's supralinear summation of EPSPs [49] represents a simple biophysical mechanism for achieving a long tail in the distribution of positive synaptic currents, a key element of our model. And, the control of firing threshold in Kenyon cells by global inhibitory input is well suited to endow these cells in principle with the ability to change response selectivity, for example if required by external circumstances.

### Function of selectivity reversal

Our model can reproduce the selectivity reversal seen in HVC<sub>X</sub> neurons [11]. We predict this to be a widespread phenomenon. We predict that sparsely firing neurons, when they are depolarized by constant current injections to fire densely, will display reduced or negative selectivity for their normally preferred stimulus. Similarly, we predict that densely firing neurons should lose or also reverse their selectivity while being hyperpolarized by constant current injection. These predictions apply to the mean and median selectivities in a large population (not to each individual cell) and should be relatively simple to verify using intracellular recordings. Note that these two predictions are surprising for neurons with monotonic frequency-current (F-I) curves  $f$  as in our model. We can speculate about the function of such selectivity reversal, if indeed widespread. If it were to be found in other animals and brain areas and were under volitional control, it could be used to attentionally screen the sensory environment for highly familiar stimuli (high threshold case), or to tune in on all kinds of stimuli with preference for unfamiliar ones (low threshold case).

Regarding shifts in excitatory/inhibitory balance, our model predicts that the effect on response selectivity depends on how balance shifts affect firing rates. For example, if increased inhibition leads to decreased firing rates, our model predicts increased response selectivity (for the BOS or an equivalent stimulus). In general, manipulations of excitation or inhibition within a network can lead to highly non-trivial reactions, e.g. disruption of local inhibition onto a cell can lead to lower baseline firing rate and to significant changes in firing patterns. For example, in Rosen and Mooney [50], decreased G-protein coupled inhibition led to decreased baseline firing, which is counterintuitive and may be caused by nonlinear priming effects. Our model is not able to explain such behavior, as it would need to

include a more complex neuronal model including synaptic feedback. Nevertheless, because our theory applies in the direction of firing rate changes, our model predicts increased selectivity for the BOS when removal of inhibition decreases firing rates, which is what has been observed [50].

As a model of HVC responses, our findings suggests that vocal-auditory mirrored activity in HVC has a sensory origin (activity in  $HVC_X$  neurons is mirrored in that auditory-evoked and singing-related responses in these cells are nearly identical [22]). In particular, our interpretation is that initially, HVC responses are shaped by the TUT; thereafter, HVC responses and their selectivity are further shaped by auditory feedback elicited by the BOS [21], which during early sensorimotor song development is generated by a motor pathway that excludes HVC [51]. During this developmental phase, a network forms among HVC neurons and ultimately produces adult song. A sensory origin of the HVC network would imply that motor responses in HVC neurons learn to mirror sensory responses, not vice versa (HVC neurons learn to use auditory-feedback-elicited responses as future motor outputs, rather than their learning to map auditory feedback onto the HVC neurons that were involved in generating the feedback). In other words, when mirror neurons fire during motor behavior, they do so mainly because they have developed selectivity to the stimulus preceding their firing. More specifically, mirrored activity in HVC neurons could derive from essentially one assumption: that the local HVC network tries to maximize the drive of cells at the moments at which these fire, initially driven by sensory input. Accordingly, HVC synapses would allow for cells to drive each other at time lags at which their preferred TUT/BOS auditory features occur. Such specific function could arise for example by virtue of some spike-time dependent synaptic plasticity mechanisms [52–55].

In conclusion, the architecture we have described shows that efficient coding constraints can explain the diversity of response specificity in higher sensory areas. Sparse/selective and dense/antiselective responses are at opposite extremes of the same efficient coding principle. It is possible that this link between response specificity and firing sparseness holds true also in other neural systems such as the neocortex. And, by extrapolation, our work shows that efficient coding constraints may guide the formation of sensory pathways all the way up to premotor areas, by which our work can shorten the gap between our understanding of pure sensory and pure motor codes.

## Methods

### Cochlear input

The cochlear input to the network at time  $t$  was formed by the most recent  $\tau = 50$  ms window of the (mean-subtracted) log-power sound spectrogram  $x(f, u)$ , where time  $u$  ranged from  $t - \tau$  to  $t$ , and the frequency  $f$  from 0 Hz to 11 kHz. To form these spectrograms, the sound was sampled at 22 kHz, multiplied with a Hanning window of 256 or 128 samples, and Fourier transformed, where in each frequency band we subtracted the mean spectral log-power. The frequency resolution of spectrograms was 86 or 172 Hz, respectively. The temporal offset  $\Delta t$  between subsequent cochlear inputs was 16 or 32 samples, corresponding to 0.7 and 1.5 ms, respectively. The dimension  $N_0$  of the cochlear input was 8192 for high-resolution spectrograms (86 Hz, 0.7 ms, Figures 2B and 5 to 13) and 2048 for low-resolution spectrograms (172 Hz, 1.5 ms, Figures 2A and 3). For Figure 4 the spectrum ranged from 0 kHz to 8 kHz at a resolution of 62.5 Hz, and a temporal resolution of 4 ms was used.

### Network model

The synaptic current  $y_i^t$  to neuron  $i$  ( $i = 1, \dots, N$ , where  $N$  varied from 100 to 800) was formed by convolving the cochlear input  $x(f, u)$  in time with synaptic weights  $\xi_i(f, t)$ , and by summing over frequency bands:

$$y_i^t = \sum_{v=0}^{50ms} \sum_{f=0}^{11kHz} \xi_i(f, v) x(f, t - v). \quad (1)$$

We interpret the synaptic weights  $\xi_i(f, t)$  as the STRF of cell  $i$  [56]. To simplify our notation, we vectorize the cochlear input  $x(f, u)$  as a column vector  $X^{t-\tau:t}$  and the synaptic weights  $\xi_i(f, t)$  as a row vector  $\xi_i$ , allowing us to rewrite Equation 1 as the scalar product

$$y_i^t = \xi_i \cdot X^{t-\tau:t}. \quad (2)$$

We considered two separate models for transforming synaptic currents into firing rates. In our threshold-linear or analog model, the firing rate  $r_i^t$  of cell  $i$  at time  $t$  was proportional to the total synaptic current above a firing threshold  $\theta$ ,

$$r_i^t = \max(y_i^t + \eta_i^t - \theta, 0), \quad (3)$$

where  $\eta_i^t \sim \mathcal{N}(0, k^2)$  represents a source of independent Gaussian white noise of mean zero and standard deviation  $k$ . This noise modeled fluctuating firing rates when  $k > 0$ . Note that we also simulated a second binary model, in which neurons fired at a constant rate when the total synaptic current exceeded  $\theta$ ,  $r_i^t = \Theta(y_i^t + \eta_i^t - \theta)$ , where  $\Theta(\cdot)$  is the Heaviside function,  $\Theta(x) = 1$  if  $x > 0$  and  $\Theta(x) = 0$  otherwise.

Neural firing in our model is set by a separating hyperplane of the input space  $X^{t-\tau:t}$ . Note that such a linear thresholding scheme might not be optimal for non-Gaussian stimulus ensembles [57]. We may study extensions from planar to curved decision boundaries in future work, provided suitable mathematical tools can be devised.

### Networks of two layers

We also analyzed hierarchical networks with two layers, with the goal to evaluate BOS selectivity in the second layer as a model of HVC.

The input  $\tilde{X}^t$  to the second layer was formed by a 50 ms window of (mean-subtracted) first-layer firing rates  $r^t$ , down-sampled from 0.7 ms to 2.8 ms (by averaging). Neuron  $i$  in the second layer ( $i = 1, \dots, N_2$ ) had a non-zero firing rate (threshold-linear or binary) at time  $t$  when its total synaptic current  $\tilde{y}_i^t = \tilde{\xi}_i \tilde{X}^t + \tilde{\eta}_i^t$  exceeded a threshold  $\theta_2$  (with  $\tilde{\eta}_i^t$  representing a source of Gaussian white noise of mean zero and standard deviation  $k_2$ ). The total number of second-layer neurons  $N_2$  was set to 800. Unlike in the first layer, the second-layer synaptic weights  $\tilde{\xi}_i$  cannot be interpreted as receptive field due to the dependence of second-layer responses on first-layer synaptic weights. To train the second-layer weights, we first whitened first-layer responses and then picked random 50-ms sets thereof (each with a new random noise) to iteratively update the sparseness transformation  $\tilde{W}$ . Hence, the training data for the second layer depended on the first-layer firing threshold  $\theta$ . For training of the second layer, we observed good convergence behavior when  $k = 1$  and  $k_2 = 0$  (though results did not qualitatively depend on noise amplitudes during learning).



## Synaptic weights and the cost function

The  $N \times N_0$  synaptic weight matrix  $\xi = (\xi_1; \xi_2; \dots; \xi_N)$  was given by matrix multiplication of a projection  $P$ , whitening the cochlear input, followed by a sparseness transformation  $W = (w_{mn})$ :

$$\xi = WP. \quad (4)$$

The projection  $P = \Lambda^{-1/2} \cdot E^T$  was determined by principal-component analysis, where  $\Lambda$  is the diagonal matrix of the  $N$  largest eigenvalues of the covariance matrix  $C = \langle X^{t-\tau:t}(X^{t-\tau:t})^T \rangle_t$  of the cochlear,  $E$  is the  $N \times N_0$  matrix of corresponding orthonormal eigenvectors, and  $^T$  denotes the transpose.

For a given projection  $P$ , the sparseness transformation  $W$  was computed iteratively by minimizing the following cost function for synaptic currents  $y_i^t$ :

$$W = \arg \min_W \sum_{i,t} f(y_i^t) \quad (5)$$

where

$$f(y_i^t) = (y_{i,+}^t - \theta) + \frac{c}{2} (y_{i,-}^t - y_0)^2. \quad (6)$$

In this last equation,  $c$  is the trade-off between the two summands,  $y_{i,+}^t$  are suprathreshold synaptic currents ( $y_{i,+}^t = y_i^t$  for  $y_i^t > \theta$  and  $y_{i,+}^t = 0$  otherwise), and  $y_{i,-}^t$  are subthreshold synaptic currents ( $y_{i,-}^t = y_i^t - y_{i,+}^t$ ).

This cost function contains a linear cost for suprathreshold currents ( $y_i^t > \theta$ ) and a square cost for subthreshold synaptic currents ( $y_i^t < \theta$ ). The function has two minima of zero cost at  $\theta$  and at  $y_0$ .  $y_0$  is a parameter that sets the current which elicits the minimum subthreshold cost. Changing  $y_0$  has no qualitative influence on our findings; in practice  $y_0$  can be set to zero both during learning and evaluation of the network (thus considerably simplifying the learning procedure described below). However, to achieve minimal reconstruction error as we will show, in some simulations we set  $y_0$  to the expected subthreshold synaptic current, defined by  $y_{E-} = \int_{-\infty}^{\theta} y p(y) dy / \int_{-\infty}^{\theta} p(y) dy$ , with  $p(y)$  being the probability density function of synaptic currents elicited by the training set. Hence,  $y_0$  should be regarded as a useful parameter if close to optimal reconstructions are desirable, but otherwise this parameter can be confidently neglected.

For most simulations we set  $\theta$  to zero during learning, implying that suprathreshold currents are depolarizing (positive), whereas subthreshold currents are hyperpolarizing (negative), Figure 1C. The factor  $c$  was normally set to one, but the simulations showed a high robustness even when increasing  $c$  by two orders of magnitude.

The function  $f(y_i^t)$  in Equation 6 has a discontinuity in  $\theta$ , which hinders gradient-descent algorithms. In practice, we smoothed  $f(y_i^t)$  locally at the threshold  $\theta$  using a sigmoidal function  $U_\theta(y) = \frac{1}{2} \left( \tanh\left(\frac{1}{s}(y - \theta)\right) + 1 \right)$  as follows:

$$f(y_i^t) = U_\theta(y_i^t) \cdot (y_i^t - \theta) + (1 - U_\theta(y_i^t)) \cdot \frac{c}{2} (y_i^t - y_0)^2. \quad (7)$$

The variable  $s$  is a smoothing constant that should be in the range of average sampling density in the neighborhood of the threshold (in our simulations we chose  $s = 0.01$  for  $\theta = 0$  and  $s = 1$  for  $\theta = 5$ ).

To avoid the trivial optimum  $W = 0$  in Equation 5, we set the following restriction to the (left) inverse matrix  $J = (j_{mn})$

$$J \cdot W = \mathbf{I} \quad (8)$$

$$\text{diag}(J^T J) = \mathbf{1}, \quad (9)$$

where  $\mathbf{1}$  is a vector of ones and  $\mathbf{I}$  the unity matrix. This constraint is motivated by functional/theoretical consideration and could potentially be accomplished by a host of complex homeostatic mechanisms, though such an analysis would go beyond the limits of this work. In different sets of simulations we also examined the weaker (volume-preserving) restriction  $\det(J) = 1$  and the stronger (orthonormal) restriction  $J^T J = \mathbf{I}$ , where  $\mathbf{I}$  is the identity matrix. For the entire range of thresholds  $\theta$  tested, the distributions of BOS-REV and BOS-CON selectivities were almost unchanged under these two restrictions, demonstrating robustness of our results with respect to the synaptic weight constraint. Also, similarity of our findings with findings obtained when using the Bell-Sejowski algorithm demonstrates high robustness. Note that all presented simulation used a squared projection matrix  $W$  where  $J = W^{-1}$  is the inverse.

Geometrically, the cost function penalizes subthreshold outliers more than suprathreshold ones, because the square Euclidean norm of a large vector is larger than its Manhattan norm. Furthermore, the cost of mapping an outlier to a suprathreshold synaptic current in just a single neuron is smaller than the cost of distributing the outlier among multiple neurons. This can be best seen for the special case of sparseness transformations that are rotations (i.e. orthonormal, satisfying  $\text{diag}(J^T J) = \mathbf{1}$ ), because the Manhattan norm of a vector of fixed Euclidean norm is minimized when it is parallel to one of the coordinate axes. Thus, the cost function tries to produce highly selective neurons that are strongly excited by only a very small set of inputs.

To have control over the density of STRFs (sharply tuned versus broadly tuned neurons) we run a few simulations with an additional term in the cost function, corresponding to a linear cost on absolute synaptic weights,

$$W = \arg \min_W \sum_{i,t} f(y_i^t) + c_s \frac{\sqrt{N}b}{\sqrt{N_0} \|P\|_F} \|\xi\|_1, \quad (10)$$

where  $b$  denotes the batch size (number of cochlear input samples per weight update),  $\|\cdot\|_F$  the elementwise Frobenius-norm, and  $\|\cdot\|_1$  an elementwise 1-norm and  $c_s$  is the relative weight of the new term. In all our simulation results reported in the Results Section and the figures, we applied  $c_s = 0$ , except results shown in Figures 2B and 5, where we used  $c_s = 0.2$ . In summary, for  $c_s > 0.1$  STRFs were of low density (Figure 2B) in closer resemblance to experimental data (with this choice of  $c_s > 0$  our findings about selectivity reversal still applied).

## STRF estimation using reverse correlation

We examined the question whether estimated STRFs using reverse correlation depend strongly on the firing threshold and whether they resemble the STRFs encoded in the synaptic weights. For various thresholds  $\theta$  we calculated the firing rates  $r_i^t = \max(y_i^t + \eta_i^t - \theta, 0)$  in response to birdsong stimuli and estimated the reverse-correlation STRFs  $h_i$  in terms of

$$h_i = (C_{ss} + \mu D)^{-1} C_{sr}, \quad (11)$$

where  $C_{ss}$  is the stimulus autocorrelation matrix,  $D$  the diagonal of the autocorrelation matrix,  $C_{sr}$  the crosscorrelation between the stimulus and the response, and  $\mu$  is a small factor used to regularize the inverse. Using the estimated STRFs we predicted the firing rates  $\hat{r}_i^t$  to novel stimuli. We compared predicted firing rates to firing rates elicited by the actual STRFs using the correlation coefficient  $cc$ .

### Minimal decoding error

The first term of the cost function in Equation 6 imposes a linear cost on suprathreshold synaptic currents. Because suprathreshold synaptic currents are equivalent to instantaneous firing rates at zero noise, the first term enforces firing sparseness across the population (for the training threshold). The second term imposes a quadratic cost on subthreshold synaptic currents, which is equivalent to minimizing an error bound on decoded cochlear inputs. To see this, consider the following estimate  $\hat{X}_p^t$  of whitened cochlear inputs  $X_p^t = PX^{t-\tau:t}$  from suprathreshold synaptic currents at time  $t$ :

$$\hat{X}_p^t = J(Y_+^t + Y_{E-}^t), \quad (12)$$

where  $Y_+^t$  is a time-dependent vector of suprathreshold currents  $y_{i,+}^t$  and the vector  $Y_{E-}^t$  is the vector with the expected subthreshold currents:  $y_{i,E-}^t = y_{E-}$  for  $y_i^t < \theta$  and  $y_{i,E-}^t = 0$  else.

The decoding error associated with the decoding  $\hat{X}_p^t$  is given by the mean square Euclidean norm between  $\hat{X}_p^t$  and  $X_p^t$ . This error is related to the distribution of subthreshold currents as follows:

$$\begin{aligned} \sum_i \|X_p^t - \hat{X}_p^t\|_2^2 &= \sum_i \|JY^t - J(Y_+^t + Y_{E-}^t)\|_2^2 \\ &= \sum_i (Y_-^t - Y_{E-}^t)^T J^T J (Y_-^t - Y_{E-}^t) \\ &= \sum_i (Y_-^t - Y_{E-}^t)^T (\mathbf{I} + C) (Y_-^t - Y_{E-}^t) \quad (13) \\ &\simeq \sum_i (Y_-^t - Y_{E-}^t)^T (Y_-^t - Y_{E-}^t) \\ &= \sum_{y_i^t < \theta} (y_i^t - y_{E-})^2 \end{aligned}$$

where in the first line we have used that  $X_p^t = PX^{t-\tau:t} = JY^t$ , and in the second line  $\mathbf{I}$  represents the identity matrix and  $C$  a matrix with zeros on the diagonal. The vector  $Y_-^t$  is the vector of the real subthreshold currents  $Y_-^t = Y^t - Y_+^t$ . The approximation in the second line is based on the assumption of equally distributed and mutually independent subthreshold currents:  $p(y_i, y_k) = p(y_i)p(y_k)$ . Note that the approximation in Equation 13 is exact for rotations ( $J^T J = \mathbf{I}$ ), whereas for the constraint  $\text{diag}(J^T J) = 1$ , we found the approximation to be within 4% of the reconstruction error for  $\theta = 0$ , and to be even closer for higher  $\theta$ .

The key insight is that for the training threshold  $\theta$ , the approximated reconstruction error is proportional to the subthreshold term of our cost function if we choose  $y_0 = y_{E-}$  in Equations 5 and 6. Hence, by minimizing the subthreshold term in our cost function, we minimize the approximation of the decoding error. The final term in Equation 13 shows that the reconstruction error is small when the subthreshold currents are rare ( $p(y < \theta)$  is small) and their variance is small ( $\langle (y - y_{E-})^2 \rangle_{y < \theta}$  is small).

Note that the decoding scheme defined in Equation 12 may not be globally optimal, but was motivated by our assumption that

only suprathreshold events carry meaningful information. Also, a benefit of this decoding scheme is its asymptotic robustness (for a threshold of minus infinity the decoding error vanishes).

Our algorithm minimizes an approximation of the reconstruction error, but not the reconstruction error itself. Exploratively, we have adapted the algorithm to directly minimize the reconstruction error (defined in Equations 13) for a given threshold  $\theta > 0$ . The resulting reconstruction error for BOS at the given threshold was only marginally better than with our sparse-coding algorithm. Synaptic current distributions were nearly bimodal with a first peak at zero and a second peak slightly above  $\theta$ ; interestingly, reconstructions became worse than in Figure 9E when performed using a different threshold than used during learning.

### Reconstructed Spectrograms

From the decoded whitened inputs  $\hat{X}_p^t$ , we reconstructed the spectrograms (Figure 9) using the pseudoinverse  $P^{-1} = E\Lambda^{1/2}$ . The element  $t$  of the reconstructed spectrogram was defined by

$$X_{rec}^t = E\Lambda^{1/2} \hat{X}_p^t. \quad (14)$$

From these elements, we computed the fully reconstructed spectrograms by averaging over all overlapping regions in the sequence  $X_{rec}^t, X_{rec}^{t+1}, \dots$  (average over 64 time windows).

### Algorithm for cost minimization

The constraint in Equation 8 was naturally enforced by our choice of parameterization of the inverse:

$$j_{mn} = \frac{\sin b_{mn}}{M_m |\cos b_{mn}|}, \quad (15)$$

where  $B = (b_{mn})$  is an arbitrary parameter matrix and  $M_m = \sqrt{\sum_l \tan^2 b_{ml}}$  is a row-wise normalization constant. We minimized the cost function  $F = \sum_{i,t} f(y_i^t)$  iteratively using a line search along the direction  $A = -\nabla_B F$  of steepest descent along the cost surface, which was given by:

$$\begin{aligned} a_{mn} &= - \sum_{i,t} \frac{\partial F}{\partial y_i^t} \frac{\partial y_i^t}{\partial b_{mn}} \\ &= \frac{\sin b_{mn}}{M_m^2 \cos^3 b_{mn}} \left( \frac{z_{nm}}{j_{mn}} - \frac{\sum_l z_{lm} j_{ml}}{M_m} \right), \end{aligned} \quad (16)$$

where

$$z_{nm} = \sum_{i,t} y_n^t w_{im} \left( c y_{i,-}^t + \sigma(y_{i,+}^t) \right), \quad (17)$$

for the simplified cost function (Equation 6) and for  $y_0 = 0$ .  $\sigma(\cdot)$  denotes the signum function. However, for the general, smoothed cost function (Equation 7) the calculation of  $z_{nm}$  gets more complex:

$$\begin{aligned} z_{nm} &= \sum_{i,t} y_n^t w_{im} \left( c \cdot (y_i^t - y_0) \cdot (1 - U_\theta(y_i^t)) - \frac{c}{2} \cdot V_\theta(y_i^t) \cdot (y_i^t - y_0)^2 \right. \\ &\quad \left. + U_\theta(y_i^t) + V_\theta(y_i^t) \cdot (y_i^t - \theta) \right), \end{aligned} \quad (18)$$

where  $V_\theta(y)$  is the derivative of the sigmoid function  $V_\theta(y) = \frac{dU_\theta}{dy} = \frac{1}{2s} \cosh^{-2} \left( \frac{1}{s} (y - \theta) \right)$ .

Starting with the identity matrix  $B^0 = \mathbf{I}$ , we updated the  $k^{\text{th}}$  parameter matrix  $B^k$  according to

$$B^{k+1} = B^k + \lambda A \quad (19)$$

using an optimized step size  $\lambda$  determined through a line search algorithm:

$$\lambda = \arg \min_{\lambda'} F(B + \lambda' A). \quad (20)$$

After each update of  $B^k$ , we updated the inverse  $J^k$  according to Equation 15 and the sparseness transformation  $W^k$  by inversion of  $J^k$ . In simulation in which the minimum subthreshold cost  $y_0$  was variable, we changed  $y_0$  at each update to the expected subthreshold current,  $y_0 = y_{E-}$ . During evaluation of the network we left  $y_0$  untouched.

Note that we also formulated an overcomplete version of our algorithm, in which there are more neurons than principal components (the matrix  $W$  has more rows than columns). In this case the resulting synaptic current distribution had an even heavier tail on the positive side, but results were qualitatively unchanged.

### Batch training

Our training data consisted of 46 sound files of average duration 1.7 s. Thirty-four of these files contained different renditions of a zebra finch song (BOS), twelve files contained conspecific songs (CON) from two independently raised adult zebra finches in our colony (six files each). Our results did not depend sensitively on these file numbers and which song was chosen as the BOS, as long as there was an overall predominance of BOS.

To train the synaptic weight matrices, we used a randomized batch learning scheme. First we computed the whitened cochlear inputs based on the matrix  $P$  derived from all 46 sound files. To train the sparseness transformation, we chose a batch size of a few thousand 50-ms windows. Each window in this set was drawn randomly from one of the 46 files and at a random time within the file. From this set of whitened inputs we computed the gradient matrix  $A$  (Equation 16), performed the line search (Equation 20), and updated  $y_0$  if necessary. This procedure was then repeated with different sets of random windows. Typically, the number of updates required to reach convergence was within two to ten times the number of neurons  $N$ .

### References

- Waydo S, Kraskov A, Quiñ Quiroga R, Fried I, Koch C (2006) Sparse Representation in the Human Medial Temporal Lobe. *J Neurosci* 26: 10232–10234.
- Hochstein S, Ahissar M (2002) View from the top - hierarchies and reverse hierarchies in the visual system. *Neuron* 36: 791–804(14).
- Margoliash D (1983) Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J Neurosci* 3: 1039–1057.
- Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I (2005) Invariant visual representation by single neurons in the human brain. *Nature* 435: 1102–1107.
- Olshausen BA, Field DJ (2004) Sparse coding of sensory input. *Curr Opin Neurobiol* 14: 481–487.
- Smith EC, Lewicki MS (2006) Efficient auditory coding. *Nature* 439: 978–982.
- Barlow HB (1972) Single units and sensation: a neuron doctrine for perceptual psychology? *Perception* 1: 371–394.
- Coleman MJ, Mooney R (2004) Synaptic transformations underlying highly selective auditory representations of learned birdsong. *J Neurosci* 24: 7251–7265.
- Bauer EE, Coleman MJ, Roberts TF, Roy A, Prather JF, et al. (2008) A synaptic basis for auditory-vocal integration in the songbird. *J Neurosci* 28: 1509–1522.
- Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381: 607–609.
- Mooney R (2000) Different subthreshold mechanisms underlie song selectivity in identified HVC neurons of the zebra finch. *J Neurosci* 20: 5420–5436.
- Weliky M, Kandler K, Fitzpatrick D, Katz LC (1995) Patterns of excitation and inhibition evoked by horizontal connections in visual cortex share a common relationship to orientation columns. *Neuron* 15: 541–552.
- Heiss JE, Katz Y, Ganmor E, Lampl I (2008) Shift in the balance between excitation and inhibition during sensory adaptation of S1 neurons. *J Neurosci* 28: 13320–13330.
- Moreau AW, Amar M, Le Roux N, Morel N, Fossier P (2010) Serotonergic fine-tuning of the excitation-inhibition balance in rat visual cortical networks. *Cereb Cortex* 20: 456–467.
- Crist RE, Li W, Gilbert CD (2001) Learning to see: experience and attention in primary visual cortex. *Nat Neurosci* 4: 519–525.
- Vates G, Broome B, Mello C, Nottebohm F (1996) Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches. *J Comp Neurol* 366: 613–642.
- Nottebohm F, Kelley DB, Paton JA (1982) Connections of vocal control nuclei in the canary telencephalon. *J Comp Neurol* 207: 344–357.
- Brainard MS, Doupe AJ (2000) Auditory feedback in learning and maintenance of vocal behaviour. *Nat Rev Neurosci* 1: 31–40.
- Amin N, Grace JA, Theunissen FE (2004) Neural response to birds own song and tutor song in the zebra finch field l and caudal mesopallium. *J Comp Physiol [A]* 190: 469–489.

In separate simulations we also included in the training set various numbers of files with cage noises such as wing flaps, subsongs and other natural sounds; results did not qualitatively change with the inclusion of these files.

### Response selectivity

We measured the selectivity of neural responses using the psychophysical  $d'$  measure [32]. This measure discriminates between two different stimuli  $p$  and  $q$  based on the average firing rates  $\bar{r}_p$  and  $\bar{r}_q$  they elicit in individual neurons (averages are formed over stimulus presentations and over their durations):

$$d' = \frac{2(\bar{r}_p - \bar{r}_q)}{\sqrt{\sigma_p^2 + \sigma_q^2}}. \quad (21)$$

The terms  $\sigma_p^2$  and  $\sigma_q^2$  in the denominator are the variances of mean firing rates across repeated stimulus presentations (this variance was nonzero in our simulations due to the additive noise in Equation 3). A positive  $d'$  value implies selectivity for  $p$ , whereas a negative value implies selectivity for  $q$ . Selectivity results did not depend qualitatively on the noise amplitude ( $k$  from 0.05 to 5). The effect of increasing the noise amplitude  $k$  was to bring  $d'$  values closer to zero, but without changing their signs. In Figure 11, BOS-CON selectivity was evaluated on 3 different CONs that were not part of the training set.

To evaluate the  $d'$  selectivity in the Bell-Sejnowski architecture, we defined the firing rate of neuron  $i$  by the symmetrically rectified expression  $r'_i = |y'_i + \eta| - \theta$  for  $|y'_i + \eta| > \theta$ , and  $r'_i = 0$  otherwise. All programming was done in Matlab (Mathworks Inc).

### Supporting Information

**Audio S1 Original Recording of a zebra finch song.**  
(WAV)

**Audio S2 Reconstruction of the song in Audio S1 after smart noise suppression.**  
(WAV)

### Author Contributions

Conceived and designed the experiments: RH FB. Analyzed the data: RH FB. Wrote the paper: RH FB.



20. Amin N, Doupe A, Theunissen FE (2007) Development of selectivity for natural sounds in the songbird auditory forebrain. *J Neurophysiol* 97: 3517–3531.
21. Nick T, Konishi M (2005) Neural auditory selectivity develops in parallel with song. *J Neurobiol* 62: 469–481.
22. Prather JF, Peters S, Nowicki S, Mooney R (2008) Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature* 451: 305–310.
23. Sen K, Theunissen FE, Doupe AJ (2001) Feature analysis of natural sounds in the songbird auditory forebrain. *J Neurophysiol* 86: 1445–1458.
24. Nagel K, Doupe A (2008) Organizing principles of spectro-temporal encoding in the avian primary auditory area field L. *Neuron* 58: 938–955.
25. Woolley SM, Gill PR, Fremouw T, Theunissen FE (2009) Functional groups in the avian auditory system. *J Neurosci* 29: 2780–2793.
26. Amin N, Gill P, Theunissen FE (2010) Role of the zebra finch auditory thalamus in generating complex representations for natural sounds. *J Neurophysiol* 104: 784–798.
27. Tumer E, Brainard M (2007) Performance variability enables adaptive plasticity of ‘crystallized’ adult birdsong. *Nature* 450: 1240–1244.
28. Andalman AS, Fee MS (2009) A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proc Natl Acad Sci USA* 106: 12518–12523.
29. Hromádka T, Deweese MR, Zador AM (2008) Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol* 6(1): e16.
30. Woolley SM, Hauber ME, Theunissen FE (2010) Developmental experience alters information coding in auditory midbrain and forebrain neurons. *Dev Neurobiol* 70: 235–252.
31. Griffin D, Lim J (1984) Signal estimation from modified short-time fourier transform. *Acoustics, Speech and Signal Processing, IEEE Transactions on* 32: 236–243.
32. Green DM, Swets JA (1966) *Signal Detection Theory and Psychophysics*. New York: John Wiley & Sons Ltd.
33. Lewicki MS, Arthur BJ (1996) Hierarchical Organization of Auditory Temporal Context Sensitivity. *J Neurosci* 16: 6987–6998.
34. Theunissen FE, Amin N, Shaevitz SS, Woolley SMN, Fremouw T, et al. (2004) Song Selectivity in the Song System and in the Auditory Forebrain. *Ann NY Acad Sci* 1016: 222–245.
35. Grace JA, Amin N, Singh NC, Theunissen FE (2003) Selectivity for Conspecific Song in the Zebra Finch Auditory Forebrain. *J Neurophysiol* 89: 472–487.
36. Hyvriinen A, Oja E (2000) Independent component analysis: algorithms and applications. *Neural Netw* 13: 411–430.
37. Bell AJ, Sejnowski TJ (1995) Blind separation and blind deconvolution: an information-theoretic approach. *Proc Internat Conf Acoust Speech Signal Process, Detroit* 5: 3415–3418.
38. Hyvriinen A, Oja E (1997) A Fast Fixed-Point Algorithm for Independent Component Analysis. *Neural Comp* 9: 1483–1492.
39. Hyvriinen A (1999) Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans Neural Netw* 10: 626–634.
40. Plumbley M (2003) Algorithms for non-negative independent component analysis. *IEEE Trans Neural Netw* 14(3): 534–543.
41. Hoyer PO (2002) Non-negative sparse coding. *Neural Networks for Signal Processing XII*: 557–565.
42. Lee DD, Seung SH (1999) Learning the parts of objects by non-negative matrix factorization. *Nature* 401: 788–791.
43. Margoliash D (1986) Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. *J Neurosci* 6: 1643–1661.
44. Gentner TQ, Margoliash D (2003) Neuronal populations and single cells representing learned auditory objects. *Nature* 424: 669–674.
45. Greene G, Barrett DG, Sen K, Houghton C (2009) Sparse coding of birdsong and receptive field structure in songbirds. *Network* 20: 162–177.
46. Davison IG, Katz LC (2007) Sparse and selective odor coding by mitral/tufted neurons in the main olfactory bulb. *J Neurosci* 27: 2091–2101.
47. Laurent G (2002) Olfactory network dynamics and the coding of multidimensional signals. *Nat Rev Neurosci* 3: 884–895.
48. Perez-Orive J, Mazor O, Turner GC, Cassenaer S, Wilson RI, et al. (2002) Oscillations and sparsening of odor representations in the mushroom body. *Science* 297: 359–365.
49. Perez-Orive J, Bazhenov M, Laurent G (2004) Intrinsic and circuit properties favor coincidence detection for decoding oscillatory input. *J Neurosci* 24: 6037–6047.
50. Rosen MJ, Mooney R (2003) Inhibitory and excitatory mechanisms underlying auditory responses to learned vocalizations in the songbird nucleus HVC. *Neuron* 39: 177–194.
51. Aronov D, Andalman AS, Fee MS (2008) A specialized forebrain circuit for vocal babbling in the juvenile songbird. *Science* 320: 630–634.
52. Bi G, Poo M (1998) Synaptic modification in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J Neurosci* 18: 10464–10472.
53. Jun JK, Jin DZ (2007) Development of neural circuitry for precise temporal sequences through spontaneous activity, axon remodeling, and synaptic plasticity. *PLoS ONE* 2: e723.
54. Fiete IR, Senn W, Wang CZ, Hahnloser RH (2010) Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. *Neuron* 65: 563–576.
55. D’Souza P, Liu SC, Hahnloser RH (2010) Perceptron learning rule derived from spike-frequency adaptation and spike-time-dependent plasticity. *Proc Natl Acad Sci USA* 107: 4722–4727.
56. Theunissen FE, Sen K, Doupe AJ (2000) Spectral-Temporal Receptive Fields of Nonlinear Auditory Neurons Obtained Using Natural Sounds. *J Neurosci* 20: 2315–2331.
57. Sharpee T, Bialek W (2007) Neural decision boundaries for maximal information transmission. *PLoS ONE* 2: e646.